DOCUMENT RESUME

ED 051 864 LI 002 924

AUTHOR	Vaswani, P. K. T.; Cameron, J. B.
TITLE	The National Physical Laboratory Experiments in Statistical Word Associations and Their Use in Document Indexing And Retrieval.
INSTITUTION	National Physical Lab., Teddington (England).
REPORT NO	COM-SCI-42
PUB DATE	Apr 70
NOTE	171p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS	*Cluster Grouping, *Indexes (Locaters), *Indexing, *Information Retrieval, *Search Strategies, Word Frequency

ABSTRACT

The experiments involved 11,571 abstracts (with titles), 1,000 key-word stems and 93 search requests. Measures of word association are derived in several ways from the numbers of documents in which two given words co-occur, and measures of similarity from the numbers of words associated with both. Word clusters with different degrees of overlap are derived from the resulting networks of word connections for use as document descriptors. All are employed in retrieval and their performance analyzed. Two new measures, sensitivity and coverage, reflect the variation in a strategy's performance from request to request. The best strategy depends on the user's requirements. For a single strategy key-words are simplest but the quantities of output are erratic and may usefully be controlled according to word associations. If two strategies can be used key-words alone may be followed by associations, yielding in a similar output quantity 30% more relevant documents. The corresponding use of clusters is marginally better but unlikely to justify its extra cost. (Author)

COM. SCI. 42

April 1970

NATIONAL PHYSICAL LABORATORY

THE NATIONAL PHYSICAL LABORATORY EXPERIMENTS IN STATISTICAL WORD ASSOCIATIONS AND THEIR USE IN DOCUMENT INDEXING AND RETRIEVAL

by

Dr. P.K.T. Vaswani and J.B. Cameron

COMPUTER SCIENCE DIVISION

Summary

The experiments involved 11,571 abstracts (with titles), 1,000 key-word stems and 93 search requests. Measures of word association are derived in several ways from the numbers of documents in which two given words co-occur, and measures of similarity from the numbers of words associated with both. Word clusters with different degrees of overlap are derived from the resulting networks of word connections for use as document descriptors. All are employed in retrieval and their performance analysed.

Two new measures, sensitivity and coverage, reflect the variation in a strategy's performance from request to request. The best strategy depends on the user's requirements. For a single strategy key-words are simplest but the quantities of output are erratic and may usefully be controlled according to word associations. If two strategies can be used key-words alone may be followed by associations, yielding in a similar output quantity 30% more relevant documents. The corresponding use of clusters is marginally better but unlikely to justify its extra cost.

Contents

PART 1

I Statement of Problem and Outline of Approach.

.1 Scope of the NPL experiments
.2 The retrieval problem
.3 Comparison of fixed classification schemes and key-word based systems
.4 Advantages of mechanical systems based upon key-words
.5 Systems based upon simple comparison of key-words
.6 The need to search under related terms
.7 Semantic relationships
.8 Structural relationships
.9 What use should be made of relational information for document retrieval?
.10 Statistical word associations
.11 Word clusters
.12 Scale and subject area of experiments
.13 Relevance and coordination
.14 Output and evaluation
.15 Recall.

PART 2 : WORD ASSOCIATION AND CLUSTERING

II Preparation and Machine Input of Texts

.1 Choice of texts and subject matter
.2 Unit of text
.3 Selection of key-words
.4 Conflation of word forms
.5 The stem dictionary
.6 Word truncation
.7 Key-punching of texts
.8 Representation of text in the computer store
.9 Dictionary organization and look-up
.10 Word-abstract incidence matrix

III Derivation of Statistical Associations

.1 Scale of experiment - what corpus size?
.2 Variation of word co-occurrence frequency with corpus size
.3 The ACE computer
.4 Matrix of word co-occurrence frequencies
.5 Degree of association and statistical confidence
.6 Choice of association factor
.7 Thresholding the association factor
.8 Comparison of real and random abstracts
.9 Computation of similarity coefficients
.10 Word-connection matrices, G3 and G4.

IV Generation of Word Clusters

.1 Graph representation of word connections
.2 Properties of clusters
.3 Completeness of clusters
.4 Fragmentation of a graph by random removal of lines
.5 Meetham's method of clustering, based upon removal of lines not included in increasing numbers of triangles
.6 Cluster emphasis.

PART 3 : RETRIEVAL EXPERIMENTS

V General Considerations

.1 Brief description of KDF9
.2 Search requests
.3 Form of output
.4 Batch processing
.5 Possible variables
.6 Quantity of output and cut-off
.7 Sensitivity of a strategy to changes in the cut-off
.8 The harvest file.

VI Indexing and Retrieval Strategies

VI.1 Description of strategies

.1 Introduction
.2 Determination of cut-off level
.3 Basic key-word stem strategy
.4 Simple use of word clusters as descriptors
.5 Use of clusters as descriptors with un-weighted representation
.6 Use of clusters as descriptors with weighted representation
.7 Word-stem comparison with expanded abstracts
.8 Word-stem comparison with expanded abstracts and requests
.9 Combined strategy - run 9
.10 Combined strategy - run 28
.11 Obligatory request words
.12 Automatic weighting of request key-word stems.

VI.2 Discussion of strategies

.1 Methods of using word associations
.2 Discussion of our use of word clusters
.3 Comparison with random superimposed coding
.4 Representation
.5 Direct use of word associations
.6 Obligatory request words.

VII Evaluation Procedure

.1 Relevance
.2 Precision
.3 Operational and evaluation parameters: K and K'
.4 Recall estimates
.5 Totals and averages.

VIII Discussion of Performance Figures

.1 General comments
.2 Relative performance of standard strategies
.3 Run 13 with a second strategy compared with Run 13 at higher cut-off
.4 Generality of a request
.5 Underlining (55 requests).

IX Conclusions

.1 Evaluation criteria
.2 Performance of key-word stem searches
.3 Advantages of associative strategies
.4 Best associative strategies
.5 Effects of cluster overlap and representation
.6 Effect of obligatory request words
.7 Effect of request generality
.8 Choice of strategy
.9 Application to on-line, interactive systems.

Acknowledgements

APPENDICES

A1 Word-Stem Dictionary and Frequencies

A2 4-Zone Code

A3 Letter Frequencies

A4 Word-Frequency Distributions
Word-Pair Frequency Distributions

A5 Binary Word-Connection Matrix G3

A6 Selection from Cluster Set C3.

B1 Complete List of Requests

B2 Numbers and Frequency Distribution of Key-Words

B3 Subsets of Requests

B4 Summary Description of Strategies

B5 Strategy Features

B6 Computer Implementation of Indexing and Retrieval Operations. Evaluation Programs and Harvest File

B7 Performance of Successive Strata in Run 13 (Key-Word Stems)

B8 Performance of Typical Strategies for Varying Cut-off K (Runs 13,28,14,25)

B9 93 Requests: Numbers of Relevant Documents

B9a 93 Requests: Numbers of New Relevant Documents

B10 34 Requests Retrieving 0-4 Relevant Documents in Run 13

B11 34 Requests: Numbers of Documents

B12 Request Sets of Increasing Generality: Numbers of Documents

B13 55 Underlined Requests: Numbers of Documents

B14 The Approximate Value of K'/K

B15 Words Accepted by Assessors in Place of Key-Words

B16 Sample Page of Subject Indexes

B17 Glossary.
Bibliography

Addenda

p.53 VII.4 after line 12 insert

We have two versions of the harvest file on which document assessments are stored. On the standard harvest file HKG50 there are 2020 known relevant documents, on the full harvest file HAR00 there are 2090.

On HKG50 the output corresponds strictly to the K values yielding $K' \simeq 50$ for the 93 requests.

On HAR00 assessments from all sources were retained in case the abstracts should be retrieved again in new runs, or in old runs at higher K'.

These extra sources were: Runs 6,9,13,20 at $K' \simeq 51,52,51,52$ respectively (5 relevant); Run 26 (42); Run 27 (24), the latter pair having 1 in common. The K' mentioned come from the slightly higher K, namely 50,50,72,74, needed to give $K' \simeq 50$ at a time when requests 1,6,46,66 had not yet been cancelled. In the other standard runs K' was not altered within the nearest integer.

The requests benefited somewhat unevenly, but the net result is that both overall and average known recall figures derived from HAR00 are lower by about 3½%.

In the report Appendix B8 is derived from HAR00; all other known recalls as on pp.53,54,Appendix B3(v),Appendix B9b are related to the standard HKG50.

Tables similar to Appendix B8 based on HKG50 are available for all runs except 24,26,27 up to $K' \simeq 50$.

Our main use of recall is to obtain the estimates of 80% on p.54: it does not enter directly into our evaluation and is tabulated for rough comparison with other experiments.

after Appendix B9a insert

APPENDIX B9b

93 Requests: Average Output $K' \simeq 50$

Run	Relevant	Irrelevant	Total	Overall % Precision	Overall % Known Recall
KWS 13	991	3637	4628	21.41	49.06
AWKWS 22	898	3720	4618	19.45	44.46
ARM 16	782	3835	4617	16.94	38.71
MCS01 14	873	3870	4743	18.41	43.22
MCS11 20	857	3667	4524	18.94	42.43
RJR 19	939	3764	4703	19.97	46.49
ARMSR 17	785	3877	4662	16.84	38.86
SR14 6	753	3852	4605	16.35	37.28
PDR14 11	861	3825	4686	18.37	42.62
EAG3 15	828	3811	4639	17.85	40.99
EAG4 18	835	3764	4599	18.16	41.34
EARG4 23	914	3793	4707	19.42	45.25
13T14 9	966	3729	4695	20.58	47.82
13W14 28	991	*3689	4680	21.34	49.06
U14 21	902	3759	4661	19.35	44.65

*includes 36 not assessed

The figure used for total known recall is 2020 relevant (c.f. VII.4.iii).

NPL Report Com Sci 42

Corrigenda and Addenda

Corrigenda

p.54 lines 3-6 read

Request 10	11	53	11	75
38	4	1	4	9
91	9	5	9	23
	24	59	24	107

Appendix B4 second page line 3 delete 6.

20 July 1970

PART 1 : Statement of Problem and Outline of Approach

I.1 Scope of the NPL experiment

During the last two decades there has been an increasing awareness of the inability of fixed classifications, such as the Universal Decimal Classification (UDC), to cope with Document Retrieval, that is the provision of a list of references in response to a request for information on a specified subject. At the same time, methods of indexing and retrieval based more directly upon the vocabulary of documents and requests have grown in popularity and there has been a marked trend towards the use of such systems by many of the more progressive technical libraries.

Document, Reference, or Text Retrieval must be distinguished from Fact or Data Retrieval. A request for articles on

'Radio Waves from the planet Jupiter'

differs in kind from the queries

'Which companies worth over a million pounds attempted takeovers in 1968?'

'What is the Specific Heat of Copper?'

In this report, we are concerned only with the former.

The work reported here, which began at NPL in 1961 and was recently brought to an end, has been concerned with the improvement of key-word techniques in purely mechanical indexing and retrieval systems. Its objectives were

(i) to develop methods of clustering words on the basis of specially computed statistical measures of association between word pairs, and

(ii) to explore and evaluate ways of employing these clusters and associations to improve performance, especially in the ability to recall relevant material.

We assume that natural language texts are in machine-readable form.

I.2 The retrieval problem

Given a collection of documents and a request on some specified topic, the problem is to chose criteria which will select the documents likely to answer the request; will select them singly, or in groups, in order of likelihood; and will optimise the system performance as judged by the users.

Thus any retrieval system must

(i) subdivide the collection, preferably into mutually exclusive subsets, and

(ii) order these subsets of documents for presentation to the requestor.

I.3 Comparison of fixed classification schemes and key-word based systems

The view of the retrieval process taken above clarifies some of the main distinctions between hierarchical (tree-like) classification schemes, such as the UDC, and key-word based retrieval systems.

In the case of a hierarchical classification scheme:-

(i) The system provides a single, fixed, subdivision of the document collection and a fixed ordering of the documents.

(ii) The system itself does not specify the order in which the document or document subsets should be considered when a search is conducted. The requestor must decide this himself. Cross-references and 'see also' entries are all the system provides to aid him in accomplishing this difficult task.

In contrast, in the case of key-word based systems:-

(i) A special subdivision of the document collection is provided for each search request.

(ii) The document subsets are ordered according to some criteria for assessing the likelihood that a document answers the request. In the simplest case, the criterion is taken as the number of key-words in the request that were also assigned to a document to index it. The collection is then subdivided so that all the documents in one subset score the same number of points and the subsets are ordered according to these scores.

There is no reason to suppose that a unique arrangement of the document collection into subsets can be found that would be best, or even good, for all possible requests. In fact, this seems highly unlikely. Yet that is an assumption underlying any classification scheme like the UDC. The chief benefit of the UDC is to decide the arrangement of the books on the shelves. In this respect key-word based systems are superior in providing a tailor-made subdivision for each request. They also have the great advantage of producing an ordering of the document subsets for presentation to the requestor. As will be seen later, some of the more elaborate key-word systems can produce a subdivision of the document collection into a very large number of subsets, producing quite a fine ordering or partial ordering of the collection. This enables the requestor to terminate his search when he desires, giving him considerable control over the quantity of output he receives from the system.

I.4 Advantages of mechanical systems based upon key-words

(i) Being mechanical, they do not suffer from human indexing inconsistencies, as do the more traditional systems such as the UDC.

(ii) The key-words may be combined as and when necessary in order to describe the subject content of, or to index, a document, or formulate a request for information. Thus, a document describing a piece of electronic equipment used for measuring

the human heart-rate could have been indexed by such words as 'medical', 'heart', 'rate', 'electronics', 'instrument', 'measurement', even before the development of the 'medical electronics' field had been anticipated, and even if those words had never before been brought together within the system. By contrast the UDC, for example, sets out to anticipate all subject areas of interest. Unpredicted areas covering more than one discipline cannot be accommodated without considerable, and in practice continual, revision of the scheme.

(iii) Being mechanical and therefore amenable to computer handling, they boast the advantages associated with modern computing machines. These are, principally, high speed and reliability and comparative ease of repeatability. At present texts are key-punched by hand from the printed copy, but more and more material is being put into machine-readable form even before printing.

(iv) The processing speeds possible with computers make it practicable to consider procedures of a complexity that would otherwise render them quite unfeasible. This is a major advantage of such systems, and it creates a vast range of potential modes of operation still largely unexplored.

Without the aid of computers it would not be sensible to consider the techniques described in this report, as the manual processing involved in the word association and word clustering stages would be quite prohibitive.

I.5 Systems based upon simple comparison of key words

The simplest systems operate by comparing a set of key-words used in formulating a search request with sets used to index documents in the collection. There are two common ways of choosing key-words when indexing: derived indexing, in which words are selected from the title or abstract and, perhaps, from the full document text, and assignment indexing, in which the words used to index a document are decided by other means, most often according to the intuitive judgement of human indexers. The usual practice is to retrieve first any documents indexed by at least all of the words used in formulating the request. If insufficient output is produced, or if the searcher thinks that not all relevant material has been gathered in this way, it is customary to relax the request formulation and to retrieve anything indexed by at least all but one of the request words. If the request consists of more than two words it can be further relaxed in this way to obtain still more output. A slightly more elaborate procedure allows the request to be formulated as a Boolean function (involving logical connectives such as AND, OR, NOT) of a set of key-words.

Control of the indexing language sometimes involves no more than recognition of synonyms and perhaps the use of some form of 'user's dictionary' or 'scope notes' giving guidance on the choice of words in an effort to achieve a modicum of indexing consistency. Key-word stems may be used in place of key words, especially when they are chosen from the text itself. This means conflating or identifying morphological variants such as 'act', 'acts', 'actor', 'acted', 'active'. This can be done mechanically to a satisfactory degree, for example by matching with as long a stem as possible in the machine-held dictionary (See II.9).

I.6 The need to search under related terms

The richness and flexibility of natural languages are such that almost any search request can be expressed in numerous alternative ways. Different people will have different ways of asking for essentially the same thing and, indeed, any individual might express the same request in different ways on different occasions. For example, one would be asking for virtually the same thing by requesting information on

'transistor phase-splitting circuits'

and on

'driving circuits for transistor push-pull amplifiers.'

A simple system of the type discussed in the previous section would not recognize this fact and, finding that the requests apparently have little in common, would process them accordingly.

Another factor is that very few requests are such that all relevant documents would be retrieved if attention were restricted to those bearing a close parallel to the form of the request. Take Bar-Hillel's example of a search request for documents dealing with

'diseases of animals in South America.'

One can envisage documents with such titles as

'Bacteria living in dogs'

and

'The life cycle of insect X'

which may well be relevant and would certainly be worthy of consideration.

This is not a matter of allowing for alternative expressions of the same idea, but one of extending the search to cover subject areas related to that of the request. Once again the recall of relevant material should be improved by searching under terms related to those in the request.

I.7 Semantic relationships

Many kinds of semantic relationship can exist between words or phrases whether within one text or in two comparable texts. Such relations include synonymy, morphological variation, attribution, cause-effect, part - whole; more broadly, similarity of function, even physical proximity where the connection is frequent enough. Almost any association can attach itself to the meanings of words over a period of time, or conversely, may decay or vanish. The strength of a semantic association may also vary from library to library. Some associations apply only within particular documents; others to a subject area of the particular library, or to the library as a whole; others more widely still. The indexing procedure and/or the search strategy should take this into account. Thus with the example given above, documents should be considered for retrieval if indexed by such words as 'bacteria' and 'insect'.

Under such circumstances, the retrieval strategy must decide the relative importance of documents indexed by request words, by semantically related words and by combinations of both.

I.8 Structural relationships

The underlying structure of a text can be described in terms of relationships between words or other components of the text. Any transitive verb, in so far as it indicates some specific sort of connection between a subject and an object, constitutes a structural relationship. All the classical syntactic relationships are structural in nature. Structural relationships are exhibited, for example, by the qualification of a verb or adjective by an adverb, or of a noun by an adjective. Again, in describing the content of technical documents, the 'effect (or action) of A upon B' might be used as a structural relationship. In this case A and B might be chemical elements or compounds, machine components, processes, environmental conditions, etc.

Consider the following title:

Effect of transverse field on switching rates of magnetic-core storage systems.

Examining some of the syntactic relationships we note that 'storage' qualifies 'system', 'magnetic' qualifies 'core', 'magnetic-core' qualifies 'storage system', 'switching rates' pertains to 'magnetic-core storage system', 'transverse' qualifies 'field', etc.

These structural relationships, unlike semantic ones, have no existence outside the texts in which they occur; they are properties of particular texts rather than properties of large corpora.

In the context of document retrieval systems it is still far from clear what use, if any, can be made of structural relationships existing in the documents or in requests. Is it helpful to preserve the structural information as far as possible when indexing? This is done in the SYNTOL system [1] and also in the intricate relational indexing system developed by Farradane [2]. On a much lower level many systems employ <u>links</u> to indicate the existence of relationships at the indexing stage: words, phrases, etc., that are interrelated are tagged and pairs of common tags indicate entities which are related, without defining the sort of relationship.

If structural relationships are specified when indexing documents and only those documents retrieved which show a sufficiently good correspondence with the structure of a request (it being automatically assumed that the words used to index the retrieved documents correspond satisfactorily with those of the request), then the chances are very high that documents of likely interest will be missed. Once a request is transformed or matched against a document using semantically related words, the structure is likely to lose its parallel, particularly, of course, if the matching words are spread over several sentences in the document.

I.9 What use should be made of relational information for document retrieval?

From the discussion in the previous section there seems little case for the introduction of structural or syntactic relationships. On the whole, such information is far too specific in nature to be useful if one is concerned in

improving the recall of relevant documents. If one's prime interest were in minimizing the recall of irrelevant documents, or if one were setting up a fact-retrieval system, different considerations would apply.

On the semantic side synonymy and near-synonymy are not the only things of interest, all other relationships are potentially useful as pointers to words, additional to those used in a request, that should be considered in the retrieval strategy. Our prime concern is to improve retrieval systems from the point of view of their ability to recall as much relevant material as possible. That being so, the most valuable information is that specifying which words are semantically related. There is little further advantage in distinguishing between the various kinds of relationship occurring.

Cor eration shows that the necessary information about semantically related ts is not readily available. The task of investigating and recording all the relationships for a large technical vocabulary is enormous. Furthermore, the job would have to be repeated from time to time because the situation is dynamic, new relationships being formed and old ones dying out as subjects develop. To some extent the relationships existing depend upon the interests and subject coverage of each particular library. For these reasons, and in the interests of greater objectivity, it would be a tremendous advantage if the identification of semantic relationships could be mechanized. A possible method of doing this, which has been tried experimentally, is discusse in the next section.

I.10 Statistical word association

Suppose that a large set of texts is analysed statistically, observations being made of the total number of distinct words, of the number of sentences containing each word and the number of sentences containing each possible pair of distinct words. From a knowledge of the number of sentences in the full set of texts and of the number of sentences containing each word, and making the assumption that words are statistically independent in their occurrence in texts, it is simple to calculate the expected number of sentences containing any given pair of words. If the number of sentences in which two words are actually observed to occur does not differ too greatly from the calculated expected number, then it is reasonable to suppose that the assumed statistical independence of the given words is verified. If, on the other hand, these observed and expected quantities are significantly different, then the assumption has not been verified and the words in question must be assumed to be statistically associated. Techniques exist for deciding what is significant in this respect and for obtaining a quantitative measure of the degree of statistical association between two words.

Suppose the sample of texts is so large that the measured statistical associations may be taken as representing properties of the language, rather than peculiarities of the particular sample of texts (at least within a given subject area if not in general). It is then reasonable to hypothesize some, albeit unidentified, semantic relationship whenever a significant statistical association is found. The measure of association being statistical, there is always a chance that a high association that looks very significant has, in fact, arisen fortuitously. Steps must be taken to control this. These considerations are fully dealt with in section III, where we define and measure an association factor for any pair of words. We cannot assume directly that, because two words frequently co-occur within texts, therefore one is a good substitute for the other.

Accordingly, for retrieval purposes, we develop a similarity coefficient derived, again mechanically, from the association factor. This measures the tendency of two given words to be found in similar company.

I.11 Word clusters

Having established a network of statistical connections between a set of words by a technique of the sort just described, it is interesting to investigate whether there is a tendency, by virtue of these interconnections, for the words to cluster. Assuming fairly restrictive rules of cluster formation, so that the degree of interconnection within clusters is relatively high, identification of such clusters offers the following advantages:-

(i) A useful relationship for retrieval purposes may be hypothesized between any pair of words in a cluster. This provides a way of predicting useful connections between words in many cases where they do not result directly from the initial statistical analysis.

(ii) Instead of indexing documents in terms of individual words or phrases, each cluster of closely associated words could be used as a descriptor* to be assigned to documents to index them and to be used for formulating requests. This method of utilizing the word associations, if effective, is economical in terms of storage space required for indexing a large collection. Further economy is achieved by not having to store explicitly all the separate word associations.

I.12 Scale and subject area of experiments

A collection of some 12,000 abstracts of papers in electronics, computers, physics and geophysics was used as the corpus for the statistical analysis and derivation of word associations. The same collection formed the document base for later retrieval runs in which various indexing and search strategies and several different sets of word clusters were tested. A dictionary containing 1,000 key-word stems was compiled on the basis of a sample of the abstracts. These are the only words, taken separately or in clusters, used in the experiments for indexing and searching.

Each retrieval run involved searches for 93 requests, 7 of an original set of 100 having been cancelled. A variety of strategies involving the use of five different sets of word clusters, and some strategies employing the statistical associations directly, no clusters being involved, have been tested and compared in a total of 14 main retrieval runs.

I.13 Relevance and coordination

All evaluation of performance is based on subjective considerations of the relevance of retrieved material by the 20 people who supplied the search requests, each assessing his own as relevant or not. Within this report we use the word relevance to refer to this subjective judgment by the user, and not to any decision by the machine. The machine as we have said earlier

---

*By descriptor is meant any key-word, phrase, word cluster, decimal classification number, author's name or any other entity assigned to documents for indexing purposes, or used to formulate search requests.

(I.2) puts the documents into subsets for presentation in a particular order. Each strategy has its own scoring system for this purpose. The score measures in some definite way the extent to which the key-words in an abstract match those in the request, and we refer to it as the coordination. Then each subset contains all documents with a given coordination, whether relevant or not. The machine has to decide how many subsets to output, beginning with the highest coordination, and for this it requires further information from the user.

I.14 Output and evaluation

A customer with a request must specify approximately the size of output he requires, for various reasons. First, there is likely to be a limit on the number of documents he is willing to scan. Secondly, the cost of an operating system is closely related to the quantity of output [3]. Other methods of specification such as a desired coordination level are likely to occasion very wide and therefore costly or inconvenient variations. It is not easy to foretell, for example, the effect of asking for all documents with a particular number of key-words present. Accordingly, strategies are evaluated by comparing the numbers of relevant documents in a given size of output, and also by the relevant documents retrieved by one strategy and not by another.

I.15 Recall

In experiments where the whole library has been evaluated, that is every document compared with every request, it has been common practice to generate output for each request till a certain proportion of its relevant documents have been retrieved (recall ratio) and then to measure what proportion these are of the output (precision ratio). One chief reason for doing this has been to allow for requests with few and with many documents in the collection. It does not however correspond very closely to the actual business of running a system, where the relevant documents are not known beforehand. In our own experiment this would require 11,571 times 93, say one million relevance assessments, which was beyond our resources. In fact about 17,000 were made by our 20 requestors. To estimate the actual numbers of relevant documents by sampling from part only of the library would have been ineffective, since the average is only 21.7 per request. However, we felt that with 14 different strategies we were likely to cover most of the relevant documents. An almost exhaustive search with 3 requests and another using subject indexes with 12 requests suggested that in fact we had already found about 80% of the relevant documents in the collection. We tabulate a known recall ratio corresponding to the numbers of known relevant in a given output.

PART 2 : Word Association and Clustering

II Preparation and Machine Input of Texts

II.1 Choice of texts and subject matter

The principal use of the texts in the first phase is word and word-pair counting. If the statistical associations produced by analysing the texts are to be as meaningful and useful as possible, texts should be used which do not contain a high proportion of repetition or redundant 'padding' material. Abstracts have the desired characteristics and that was the form of text selected for these experiments.

The abstracts were compiled at the Radio Research Station (now Radio and Space Research Station) at Slough, England. They were published over the period 1953-62 in a journal which, during that time, has been called 'Wireless Engineer', 'Electronic and Radio Engineer' and 'Electronic Technology'. The same abstracts were published concurrently in New York in the Proceedings of the Institute of Radio Engineers (now the Proceedings of the Institute of Electrical and Electronics Engineers).

Out of some twenty categories under which these abstracts appeared the following five were selected for our experiments:

1. Automatic computers
2. Circuits and circuit elements
3. General physics
4. Geophysical and extraterrestrial phenomena
5. Subsidiary apparatus.

These subjects were chosen partly because we have experts at NPL in electronics and computers who could assist us. At the same time it was thought that, from the retrieval point-of-view, these subject areas would provide most of the problems and difficulties encountered in other scientific and technical areas. Geophysical material was included so that the collection should not be too narrow in scope. We have been very grateful for the willing cooperation of the Radio and Space Research Station in providing us with the advice and assistance of subject experts in this field.

It should be pointed out that the scope of the abstracts is not as wide as the category headings would suggest. Abstracts appear under these headings only for papers having some pertinance to radio-communications.

II.2 Unit of text

In studying word co-occurrence in texts for the purpose of computing statistical association measures, what should constitute co-occurrence? Juxtaposition of two words might be taken as the criterion. Alternatively, two words may be said to co-occur if they appear in a text within some broader, but well defined, context such as a span of a specified number of words, a sentence, a paragraph, a page or a complete paper. In any case it is necessary to work with some such unit of text and to define word co-occurrence in terms of it. The disadvantages of adopting too small a unit are that many useful associations might thereby be missed, and secondly that an inordinate amount of text must be analysed. For example, if word juxtaposition were taken as the basis for co-occurrence an astronomical quantity of text would have to be scanned in order to uncover a significant

proportion of useful associations. The main penalty for working with too large a unit of text is that there would then be a high chance that any associations produced would not be very meaningful. There is probably an optimum unit of text for any proposed method of analysis. A considerable amount of experimentation would be required to establish this optimum. However, the point was not thought to be sufficiently critical to justify this and the choice was made less objectively.

Each title and abstract was regarded as a unit of text, these units containing on average a total of about 33 words. Repeated co-occurrence of two words in abstracts of papers on fairly well defined topics should be a reasonable indicator of the existence of a useful relationship between them. Many of the fortuitous relationships found in this way should be eliminated by rejecting the weaker statistical associations.

II.3 Selection of key-words

This is based upon a study of a sample (taken from the corpus) of 1,648 abstracts. A listing was produced of all the distinct words occurring in this sample together with the frequency of occurrence of (i.e. the number of abstracts containing) each. This list was studied very carefully by three people, two of them being fairly familiar with the subject matter, who decided intuitively which words to retain in the system as key-words, all others being excluded from further consideration.

II.4 Conflation of word forms

Where it was thought to be appropriate no distinction was made between different form of the same word. For example, computer, computers, compute, computing, computable, computation, etc., are not distinguished. These forms are conflated by identifying all words starting with the stem COMPUT. This is consistent with the view expressed earlier that structural details are relatively unimportant for document retrieval systems, the information discarded by conflation being mainly syntactic.

Conflation also has the following advantages:

(i) It reduces the size of the dictionary (of stems) to be stored and manipulated in the computer, and

(ii) When computing statistical association measures between word stems the frequencies and co-occurrence frequencies involved will be greater, for a given corpus, than those pertaining to the separate word forms. This enables us to work with a smaller corpus.

Another way of thinking about this is that, from the point-of-view of indexing documents and their comparison with requests, it is important that the concept of 'computation' be discussed in a document or introduced in an abstract, but it seems of relatively little interest to know whether the noun or verb is used or whether machines are discussed in singular or plural.

II.5 The stem dictionary

The system of word stems was constructed so as to maintain those distinctions between words thought to be useful for the purpose of these experiments and to suppress others. As an example, the stem ANALY appears

in the dictionary, but ANALYSER is also entered as it was decided that it might be useful to be able to distinguish such references to pieces of hardware (differential or spectral analysers) from other occurrences of the shorter stem. In a few cases word forms that would have been conflated have not been, because to do so simply by shortening the stem would permit undesired words to match the same stem. The stem CIRCL matches with circle, circling, etc., but had the stem been shortened to CIRC in order to match also with circular an undesired match would occur with the word circuit. The dictionary therefore contains CIRCULAR as a separate entry.

Having conflated terms in this way all the stems were listed and their frequency of occurrence within the 1,648 abstract sample was recorded. At this stage there were 1,324 stems. Of these, 323 occurred in the sample once only and 171 occurred twice. It was decided that most of the words occurring in the sample only once were of little value in the experiment and should be discarded. However, fourteen of them were retained as being useful words within the subject area in spite of their single occurrence in the sample.

On consideration of possible methods of organizing and handling the dictionary and other data in the computer it was found that if the number of items in the dictionary could be limited to 1,000 much more efficient use could be made of the machine and programs would be executed more rapidly than would be the case with a larger dictionary. The decision was therefore taken to discard a further fifteen of the rarer words, leaving exactly 1,000 key-word stems in the final dictionary. The complete dictionary is shown in Appendix A1.

II.6 Word truncation

It was observed during construction of the dictionary that practically all the distinctions required to be made between words could be made on the basis of the initial eight letters, or fewer, of each word. Again in the interests of economy, on being read into the computer all text words of greater length are truncated after the initial eight letters. The very small number of distinctions which consequently cannot be made seems a small sacrifice for the computer storage space saved. This is purely a matter of convenience, and although the loss of performance incurred is acceptable for the purpose of our experiments it might not be considered so if the design of an operational system were being considered.

II.7 Key-punching of texts

A typical abstract from the corpus is shown in Fig. II.7.1. The title and full text of each abstract are coded for machine input, with the following exclusions:

1) abbreviated bibliographic details (i.e. author, journal, etc.)
2) numerical data,
3) special symbols,
4) all punctuation marks other than the full stop,
5) full stops in abbreviations.

551.510.535 **2052**

Sequential E_s and Lunar Effects on the Equatorial E_s.—S. Matsushita (*J. Geomag. Geoelect.*, Sept. 1955, Vol. 7, No. 3, pp. 91–95.) Among the various types of E_s which have been observed, one shows apparent vertical movement on the ionogram, and has been termed 'sequential E_s' by investigators at the National Bureau of Standards. A study is made of this phenomenon using records from a number of stations; the subtype investigated is that involving an E_s region which first appears at a height of about 200 km in winter and 180 km in summer and then drops to normal E level, where it persists for some hours. The latitude and time distributions of the phenomenon are briefly discussed.

Fig. II.7.1. Sample abstract from corpus

This is key-punched on to 80-column cards using a standard 4-zone code (see Fig.A2.1, Appendix A2). The one spare code position is used to represent a full stop. The termination of an abstract is marked by two full stops. Each character or space occurring in the material being key-punched is coded in a separate column on a card, the first 48 columns only of each card being used for this purpose. The coding of each abstract commences on a new card and abstracts involving more than 48 characters are continued on a subsequent card or cards. Fig. II.7.2 shows the form in which the sample abstract is coded for machine input after the various exclusions have been made.

```
SEQUENTIAL E AND LUNAR EFFECTS ON THE EQUATORIAL E.
AMONG THE VARIOUS TYPES OF E WHICH HAVE BEEN OBSERVED
ONE SHOWS APPARENT VERTICAL MOVEMENT ON THE IONOGRAM
AND HAS BEEN TERMED SEQUENTIAL E BY INVESTIGATORS AT
THE NATIONAL BUREAU OF STANDARDS. A STUDY IS MADE OF
THIS PHENOMENON USING RECORDS FROM A NUMBER OF
STATIONS THE SUBTYPE INVESTIGATED IS THAT INVOLVING AN
E REGION WHICH FIRST APPEARS AT A HEIGHT OF ABOUT IN
WINTER AND IN SUMMER AND THEN DROPS TO NORMAL E LEVEL
WHERE IT PERSISTS FOR SOME HOURS. THE LATITUDE AND
TIME DISTRIBUTIONS OF THE PHENOMENON ARE BRIEFLY
DISCUSSED.
```

Fig. II.7.2. Sample abstract from corpus as key-punched.

The exclusion, during the key-punching operation, of various parts of the text was for the purpose of simplifying the task and thereby reducing its cost, and associated error rate. This applied particularly to bibliographic data. Much of this work had to be contracted out, but it was very difficult to find people with experience in this kind of work. We therefore had to make the requirements as simple as possible in order to get it done at all.

In spite of these precautions the key-punching took many months and the error rate was alarmingly high at first. In the first batch of about 1,600 abstracts, key-punched by an outside agency, about one word in twenty five was incorrectly punched. This material was prepared by girls operating ordinary hand punches, making two finger depressions to obtain the alpha-numeric

coding. Most of the remaining punching was done on machines having a typewriter style keyboard. Fortunately the error rate for the 10,000 or so abstracts following the original batch dropped to about one incorrect word in 250. Presumably this is at least partly attributable to the use of card punches with a full alphanumeric keyboard.

The latter error rate of 1 in 250 was not expected to have a pronounced effect upon the outcome of the statistical analyses and was therefore acceptable. The higher error rate in the first batch would not have been acceptable had it persisted. However, as it affected only a small proportion of the corpus those abstracts were also used without correction.

II.8 Representation of text in the computer store

Stems in the dictionary and text words read from punched cards are stored in the computer in an extremely simple coded form. Each letter, and space, is represented by a five digit binary number: the nth letter of the alphabet by n, space by the number zero. The coded representations of successive characters of a word or piece of text, taken in their natural order, are held in successive groups of five binary digits in the computer store, progressing from the more significant to the less significant end of each computer word.* To illustrate this consider the representation within the computer of the word 'machine'. These characters are first transformed into a sequence of numbers representing the position of each character in the alphabet, thus

13, 1, 3, 8, 9, 14, 5.

The binary representation in the computer is obtained simply by grouping together the corresponding five digit binary numbers:

M	A	C	H	I	N	E
01101	00001	00011	01000	01001	01110	00101

↑
more significant
end of computer word.

The text of each abstract, when read into the computer from punched cards, is stored as a continuous character string, running on from one computer word to the next and occupying as many words of computer store as are necessary. The text words are then separated by identifying the space characters, which have their own particular code. After truncating long text words by discarding all letters after the eighth, each text word is stored in a separate computer word, each of the latter being filled from the more significant end.

---

*A computer store is commonly composed of units referred to as 'words', each able to hold a string of binary digits. Each word of the Ace Computer, used in these experiments, contains 48 digits. By the more significant end of a word is meant the end at which, if the string of digits stored in the word is regarded as a binary number, the most significant digit of that number is located.

II.9 Dictionary organization and look-up

Since text words are truncated to a maximum of eight letters when they are read into the machine, no character string of greater length is ever compared with the stems in the dictionary during the look-up process. The longest stems in the dictionary therefore contain eight letters. The stems are stored one per computer word, coded in the same way as the text words. The dictionary is arranged in eight sections, according to stem length, the stems in each section being arranged in alphabetical order.

The method of coding described requires 40 binary digits to represent the longest stems. Words in the Ace computer are 48 digits long, so that each computer word containing a dictionary stem has at least eight spare digits, some of these digits are used to store a number, in the range 1-8, associated with each stem indicating the longest text words to be considered when testing for a 'match' with that stem. By definition, a match cannot exist involving a text word of more than the specified maximum length. However, since long text words are truncated after eight letters, when the number specified is eight, there is effectively no restriction on the length of text words with which the corresponding dictionary stem might match. The number indicating the maximum allowable extension of each stem is shown with the dictionary in Appendix A1.

To illustrate this system, consider the following dictionary entries (which are not adjacent):

BAND 5
BANDWIDT 8.

The first of these stems is intended to match with band and bands, no match with words of more than 5 letters being permitted. The word bandwidth is picked up by the other stem. Although it does not arise in this case, the system as defined so far could result in the occasional mismatching of a text word with a stem. Consider the following dictionary entries:

MODEL 6
MODE 5.

Since the stem MODE can be matched with words having 5 letters a mismatch could arise between this stem and the word model. However, this is avoided by accepting the longest stem with which a given text word matches. The dictionary being in sections of different stem length, this simply entails referring each word first to the section containing the longest stems, then to the others in order of stem length, and always accepting the first match for each word.

The text of an abstract having been read into the computer, before its words are referred to the dictionary they are alphabetically ordered and multiple occurrences of words are eliminated. The words in each section of the dictionary are also in alphabetical order, so all the words from the abstract are compared with the stems, and matches noted, in a single pass of the dictionary. The dictionary is stored on magnetic drums, so this procedure saves a great deal of time in terms of drum accessing.

Corresponding to each abstract processed in this way a <u>post-abstract</u> is output on punched cards containing the code numbers of all dictionary stems occurring in the abstract. Other data output at the same time include the total number of words (including repetitions and words not in the dictionary)

and the number of distinct dictionary stems in the abstract.

II.10 Word/abstract incidence matrix

This is a binary matrix (i.e. a matrix whose elements all have value either 0 or 1) having a row corresponding to each dictionary word stem and a column corresponding to each abstract (Fig. II.10.1).

		ABSTRACT					
		0	1	2	3	....	11,570
	0	0	1	0	1	....	0
	1	0	0	0	0	....	1
	2	1	1	0	1	....	0
WORD	.						
	.						
	.						
	.						
	999	0	0	0	1	....	1

(VALUES FICTITIOUS)

Fig. II.10.1. Word-abstract matrix

The complete matrix is far too large to be held in the machine, even on the magnetic drums, in its entirety. It is therefore constructed in sections, each having 1,000 rows in correspondence with the word stems and 1,536 columns, and therefore representing only a fraction of the collection of abstracts.

Before constructing a section of the incidence matrix all the elements are set to zero. The post-abstracts are then read into the computer in sequence. As each one is read in the rows of the matrix representing the word stems listed in the post-abstract are read from the drums, the element in each row corresponding to the particular abstract is set to unity and then the matrix rows thus modified are written back on the drums. After each section of the word/abstract incidence matrix is completed it is output on punched cards.

III Derivation of Statistical Associations

III.1 Scale of experiment - What corpus size?

There are two main considerations in deciding the scale of experiments of this kind. The first is that it should be large enough for the experiment to be realistic. The second is that the corpus should be sufficiently large to yield statistically significant results.

Such an experiment should be regarded as being realistic if its results and conclusions can be extrapolated and applied to the solution of problems arising in real-life retrieval situations. Some of the problems encountered in practice arise through semantic ambiguity, the interpretation of words having different meanings or connotations in different contexts and through the vague and imprecise use of words. In any experiments, therefore, the vocabulary should be sufficiently large and the subject coverage of the collection sufficiently diverse to ensure that factors such as these produce significant problems to be contended with. Results would be worthless if such difficulties were eliminated at the outset by oversimplification in setting up an experiment.

How many documents should a system provide in response to a customer's request? This question has no simple answer. It will depend upon the nature of the request, the nature of the customer's interest in the subject and upon many imponderable factors. However, most customers will have only limited time to consider the output from any system. Hence, given a vast collection and a system capable of turning up more and more relevant documents the limiting factor remains the maximum number of documents the enquirer thinks it reasonable to receive. In other words, the larger the document collection the more discriminating power required of the system. This is another disadvantage of experimenting with an unrealistically small collection.

Turning now to the subject of statistical significance, it should be noted that nothing can be inferred from a single co-occurrence of two words, and two may mean very little, regardless of the collection size. (As explained, the basic unit in these experiments is the word-stem, and 'word', used here for brevity, should be interpreted in this way). It is therefore important to ensure that the outcome of an experiment is not critically dependent upon associations derived from such very small co-occurrence frequencies, many of which may be purely fortuitous. In order to obtain larger co-occurrence frequencies in a reasonable number of cases most of the words should appear in the collection at least a dozen or so times. Rarer words are unlikely to yield the higher co-occurrence frequencies desired.
An analysis of our corpus* shows that:

10% of the dictionary stems occur in not more than 1 in 750 abstracts,
25% of the dictionary stems occur in not more than 1 in 400 abstracts, and
66% of the dictionary stems occur in not more than 1 in 100 abstracts.

These figures make it clear that, for the given subject coverage and

---

*Individual word frequencies are included in the listing of the dictionary in Appendix A1. Distributions of word frequencies, for the full corpus and for different sized subsets, are shown in Appendix A4.

vocabulary, the corpus must contain at least about 5,000 abstracts. With fewer abstracts than this many of the rarer words would yield no useful associations and would effectively be excluded from the experiment. The vocabulary, which must already be regarded as being of minimal size for a realistic experiment, would thereby be made still smaller.

Regarding the figure of 5,000 as being borderline, it was decided that the corpus for these experiments should contain a minimum of 10,000 abstracts. A total of 12,288 abstracts were finally used for the derivation of word associations and clusters, although a few of these were excluded from the document base used subsequently for retrieval purposes.

When this work began the comparable experiments of which we were aware involved very small numbers of requests and only a few hundreds of documents. (Compare [20], pp.104, 178). With perhaps one or two exceptions this situation has not altered (1969).

III.2 Variation of word co-occurrence frequency with corpus size

For convenience of handling, the abstracts were processed in batches of 1,536, that being equal to the number of bits (binary digits) in one 32-word block of memory in the ACE computer. The full corpus is composed of eight such batches. Distributions of word co-occurrence frequencies appear in Appendix A4. These are for subsets of the corpus consisting of one, two and four batches, and for all eight batches. The cumulative totals shown in Fig. III.2.1 are taken from these distributions. They indicate the number of distinct word pairs with frequency greater than each given value, excluding self-pairings. The number of such pairs possible with a 1,000 word dictionary is 499, 500.

WORD PAIR FREQUENCY	NO. OF BATCHES OF 1,536 ABSTRACTS			
	1	2	4	8
> 0	56,379	89,434	144,346	202,721
> 1	17,601	36,078	73,392	124,009
> 2	8,276	19,781	46,251	87,534
> 5	2,092	6,522	19,167	43,911
> 10	532	2,174	7,882	21,824
> 20	98	552	2,610	12,698
> 50	1	50	404	2,100
>200		1	4	98
>500				27

Fig. III.2.1 Number of word pairs with given frequency.

Of all the possible word pairs, roughly 10% occur in the single batch of abstracts analysed, rising to about 40% in the full corpus. The number of word pairs occurring with higher frequencies is of special interest and it is seen that, whereas only 98 word pairs with frequency greater than 20 occur in the single batch, the corresponding number rises dramatically to 12,698 in the full corpus. Thus from the point-of-view of a study of word co-occurrence and association, as the corpus size is increased from the 1500 abstract level its value, in terms of the number of high co-occurrence frequencies, increases far more than proportionately.

III.3 The ACE computer

All the work involved in analysing the abstracts and computing statistical measures of word pair association was executed on the ACE computer. This was the only machine of its kind, having been designed and built at NPL. Since the experimental procedure has been influenced in a number of ways by considerations of what was and what was not feasible using that machine, a brief description of it follows.

ACE was a valve-operated machine having mercury delay lines as its fast access stores. The total delay line capacity was about 800 words, including 24 long delay lines each holding 32 words. The word length was 48 bits. Each instruction occupied one machine word and contained four addresses, two specifying the location of operands, the third specifying where the result should be sent and the last pointing to the next instruction to be obeyed. The cycle time of the 32 word delay lines was one millisecond. The time taken to transfer a word from one of these to another store or to the instruction register was therefore between 32μs and 1 ms depending upon the position in the delay line of the word being transferred.

The long transfer feature of ACE was particularly useful for the analysis of word-pair frequencies. It allowed the contents of any number of words of a long delay line to be regarded as a single number or bit string. Thus the addition of two such numbers or any logical function of one or two strings of up to 1,536 bits could be executed in a single instruction. The time required to obey such an instruction was 1ms when the contents of the full delay line were involved and correspondingly less on other occasions. This facility was used extensively for binary matrix processing.

32,000 words of backing store were provided by four magnetic drums. The information on each drum was held on 256 tracks, each holding one 32-word block. The drums had moving read/write heads. The average time for a track selection involving a head movement was about 35 ms. Transfer of information between a drum track and the fast store took a further 7ms.

The number of stems in the dictionary was limited to 1,000 so that a word/abstract matrix could be stored on the drums with one track assigned to each word or matrix row, the remaining 24 drum tracks being available to hold program and other small amounts of data.

A lower level of storage was provided by six magnetic tape decks. All full-valued (i.e. non-binary) matrices of word pair frequencies, association factors, similarity coefficients, etc., were stored on magnetic tape, their requirements vastly exceeding the capabilities of the drums. Constructing and processing such matrices was a relatively slow business because of the slow speed of tape operations.

III.4 Matrix of word co-occurrence frequencies

Construction of this matrix involves applying the logical or Boolean

operation, AND, to two strings of bits. This is illustrated in Fig. III.4.1, A and B being two given strings of bits. C is the string of bits produced by applying the AND operation to A and B, and contains a '1' in every position in which both A and B contain '1'.

The first step in obtaining the co-occurrence frequency of two words is to produce the string of bits obtained by applying the AND operation to the corresponding rows (stored as strings of bits, each row on a separate drum track) of the word/abstract incidence matrix. The co-occurrence frequency of the words is then obtained simply by counting the number of 1's in this new string of bits.

As mentioned earlier, the word/abstract matrix is so large that it has to be handled in sections, only one of which can be stored on the drums at a time. The matrix is composed of eight sections, each having 1,536 columns and representing that number of abstracts.

A matrix of word co-occurrence frequencies is produced from each section of the word/abstract matrix and is stored on magnetic tape. Finally, the matrix of word co-occurrence frequencies for the full corpus is produced as the sum of these eight matrices. This was done by repeated application of a program written to sum two such matrices. When adding two matrices corresponding pairs of elements are added to obtain each element of the sum.

Bit string A:	0100011010100111001 0
Bit string B:	1001110001101010111 1
Bit string A AND B:	0000010000100010001 0

Fig. III.4.1 Performing the AND operation upon two bit strings

III.5 Degree of association and statistical confidence

We are interested in two things:

1. statistical measurement of the degree of association between two words,

2. the statistical confidence with which a degree of association can be asserted.

In the general case these are independent. For an understanding of this consider the following analogous situation. It is generally accepted on statistical grounds that an association exists between cigarette smoking and the incidence of lung cancer. The degree of association is very low, since only a very small proportion of smokers are affected in this way and by no means all lung cancer is attributable to this cause. However, because the statistics have been gathered from a vast number of case histories the evidence is overwhelming. The possibility that chance events alone account for the observed statistics and that no association exists can, for practical purposes, be ruled out. In this case we say that the association is assertable with high statistical confidence.

When trying to establish whether or not two variables are associated the usual procedure is to start with the null hypothesis that they are not and then to apply a test to see whether this hypothesis can be rejected at a suitably high confidence level. The chi-square test is often used for this purpose (see, for example, [21]). It is inadequate for our purposes, however, since it would only enable us to establish whether or not two words were associated in their use, saying nothing of the degree of association.

Ideally we would like to be in a position to say at what confidence level we could accept any given measured degree of association. This calls for rather more elaborate methods which would be costly in terms of computer time to implement.

It would be particularly important to consider confidence levels if one were working with a small sample of texts and were thereby forced to work with very small word co-occurrence frequencies. Under such circumstances there would be considerable risk that measures of high association were not significant. However, we have rejected all word pairs occurring once only in our collection and, by setting a threshold level upon our association measure, we cream off only about 10% of the word pairs occurring at least twice. From Fig. III.2.1 it is evident that many of the word pairs creamed off will have occurred twenty or more times. We feel that under these circumstances the statistical significance level of selected word pairs is not of such vital importance and that one is justified in considering only the measured degree of association. This is what we have done.

III.6 Choice of association factor

One classical measure of association (see, for example, [21]), or association factor, is obtained in this context as the ratio of the observed number of co-occurrences of two words to the expected number assuming statistical independence (i.e. no association). This may be expressed as

$$A.F.1 = \frac{N.n_{ij}}{n_i.n_j}$$

where N is the total number of abstracts, n_{ij} is the number of abstracts containing both the ith and jth words, and n_i and n_j are the respective numbers of abstracts containing the ith word and the jth word.

One interesting property of this function is its symmetry. In evaluating it for a given pair of words it is of no consequence which word is considered as the ith word and which one as the jth, the same value of association will result in either case. Although, at first sight, this may seem entirely logical, there are grounds for arguing in favour of a non-symmetric measure of association yielding, in general, two values for each pair of words. This seems sensible when one observes, for example, that given a request including the word photography, relevant documents are quite likely to contain the word film, whereas documents relevant to a request containing the word film are far less likely to contain the word photography because 'film' is used in many other contexts. Thus, if the purpose of an association measure is the selection of documents containing words associated with those in a given request, symmetry is not necessarily a desirable property.

The conditional probability, $p(w_i/w_j)$, of finding word w_i in an abstract known to contain the word w_j provides a **crude** measure of association which reflects this asymmetry in the use of words. In general the probabilities $p(w_i/w_j)$ and $p(w_j/w_i)$ assume different values.

The association factor is computed for all word pairs and the results are arranged in the form of a matrix. Quite a lot of processing requiring ready access to both the rows and columns is done with this matrix stored on magnetic drums. This presents a problem since, if the rows of the matrix are stored on successive drum tracks, then reference to the columns is slow and vice versa. For this reason we were obliged to adopt a symmetric measure of word association.

The ith column and ith row of the matrix are then identical and the need to refer to both rows and columns is removed. We wished, however, to choose as an association factor a function which would assume a high value if at least one of the two conditional probabilities for a pair of words was large. The simplest function we could imagine satisfying these requirements is that obtained by selecting the greater of the two conditional probabilities,

i.e. $$\mathrm{MAX}[p(w_i/w_j),\ p(w_j/w_i)].$$

This function is not entirely suitable because, in the case of a very frequently used word, the conditional probability of that word given another word is likely to be high for many other given words. In this case a high conditional probability does not necessarily indicate a strong association between two words. Thus, each conditional probability $p(w_i/w_j)$ should be considered in relation to the unconditional probability, $p(w_i)$, of the same word. This cannot be done by taking the ratio of the two probabilities since that yields the first association factor we considered (A_1), which does not have the desired property of assuming a high value if either of the conditional probabilties is large. We therefore followed the other obvious course and took the difference between each conditional probability and the corresponding unconditional one. Hence the function becomes

$$\mathrm{MAX}[(p(w_i/w_j)-p(w_i)),\ (p(w_j/w_i)-p(w_j))].$$

In terms of

N, the number of abstracts in the collection,
n_i, " " " " containing word w_i
n_j, " " " " " " w_j
and n_{ij}, " " " " " words w_i and w_j,

this can be expressed as:

$$\mathrm{MAX}\left[\left(\frac{n_{ij}}{n_j}-\frac{n_i}{N}\right),\left(\frac{n_{ij}}{n_i}-\frac{n_j}{N}\right)\right].$$

The fact that we were not interested in negative associations enabled us to simplify the function slightly. As mentioned earlier we wished to reject all word pairs occurring less than twice, and we decided to do this by assigning the value zero to the association factor when n_{ij} is less than two. The form in which the association factor was finally computed is:

$$\mathrm{A.F.5} = \begin{cases} \left(\dfrac{n_{ij}}{\mathrm{MIN}(n_i,n_j)}-\dfrac{\mathrm{MAX}(n_i,n_j)}{N}\right) & \text{when } n_{ij} \geq 2. \\ 0 \quad \text{when } n_{ij} < 2 \end{cases}$$

It is worth pointing out that this measure takes no account of frequency of occurrence or of co-occurrence <u>within</u> an abstract. There are two reasons for ignoring these:

(i) dealing with short abstracts the important thing is that a word is introduced at all, the frequency of use within an abstract not having much significance and being partly a matter of style, and

(ii) if techniques of this sort are ever to be of practical use the computing time required for processing word pairs (of which there might be an enormous number) must be kept to a minimum.

III.7 Thresholding the association factor

Computed as above to an accuracy of 15 bits, the association factors for all word pairs are packed, three per machine word, and stored on magnetic tape. Access to information on magnetic tape is very slow, and the necessary processing of the association factor matrix would not be practicable if attempted in this way. Instead, a threshold value is set upon the association factor, two words now being regarded as associated if their measured association is at or above this threshold. In this way a binary association matrix is produced containing a '1' to indicate each word pair association and zeros elsewhere. This matrix, being far more compact, can be held on the magnetic drums and processed in very much less time. Bearing in mind the purpose of the associations, we thought that the original matrix contained enough redundancy for this thresholding process to be quite sensible.

The problem, of course, is to decide what threshold value to set. We finally adopted the somewhat arbitrary procedure of assuming, on the basis of the proposed use of the resulting binary matrix, what would be a reasonable average number of associations per word, and setting the threshold value necessary to produce approximately that number. This is achieved by first producing a distribution of the association values in the unthresholded matrix, illustrated in Fig. III.7.1. Starting with the highest association value, cumulative totals are produced of the number of associations exceeding a progressively lower threshold value, until the desired total is approximated.

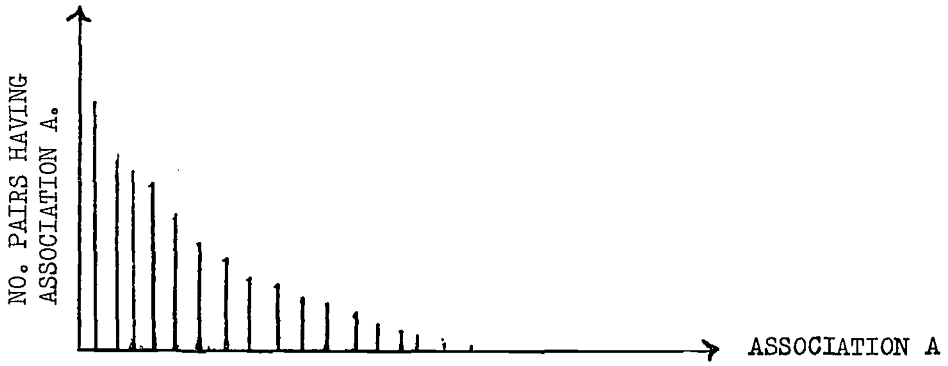


Fig. III.7.1 Distribution of association values

We initially thought that we could decide what threshold value to set by studying the distribution of association factors produced from a constructed set of random 'abstracts'. Although the exercise did not solve the thresholding problem the results obtained are very interesting and they are described in the next section.

III.8 Comparison of real and random abstracts

A set of 1,536 random 'abstracts' was constructed having the same distribution of word frequencies as a batch of real abstracts. In fact we simply produced a random word/abstract matrix having the same number of 1's in each row as the real matrix (and therefore the same word frequencies). The positioning of the 1's within the rows was determined by use of a pseudo-random number generating program.

This fictitious word/abstract matrix was processed to obtain the word pair counts and the set of association factors for all word pairs.

Distributions of word pair frequencies for the real and random abstracts appear in Fig. III.8.1. They are remarkably similar. However, the real abstracts contain significantly more high frequency (i.e. commonly occurring) word pairs than the random, as shown by the differences in the cumulative totals. The total numbers of word pairs occurring are almost identical in the two cases. Fig. III.8.2 shows the distributions of an association factor. The association factor used here is a modified form of A.F.I (section III.6), the logarithm having been introduced to compress the range of values of the function. Again there is a significant, though quite small, difference between the two distributions.

Anticipating a far greater separation of the distributions of association factors, we had thought that a suitable threshold value might be obtained by taking a value exceeded by, say, only 1% of the word pairs in the random abstracts. The hope was that such a value, used as a threshold for the real data, would reject a similar proportion of the fortuitous associations. The similarity of the distributions shows this argument to be invalid, or at any rate not to yield a useful result.

III.9 Computation of similarity coefficient

The association factor has two weaknesses with respect to the present application. The first is that we can be fairly sure that many high associations will occur by chance, in spite of the precautions we are taking. Secondly, in the short abstracts we are analyzing it is unlikely that synonyms or near synonyms will tend to co-occur, an author choosing one or the other, but having little opportunity of using both. Synonyms are therefore unlikely to be detected by means of their association factor.

The similarity coefficient was introduced in the hope that it would be better in these respects, and because we think it has far more intuitive appeal as a measure of likeness of words.

To compute the similarity coefficient between two words we compare their rows in a binary matrix of associations. In this way we are comparing the sets of words found to associate with each of the given words, and if they are similar we want the similarity coefficient to be high. This is exactly analogous to the process of computing an association factor, in which two rows of the word-abstract matrix (also binary) are compared. We have therefore employed the same function (A.F.5) used as an association factor. The reason for now referring to it as a similarity coefficient is to avoid ambiguity.

The steps involved are:

1. Produce binary association matrix from matrix of association factors, A.F.5, setting a threshold value of 5/64.

2. Compute similarity coefficients as

$$\text{S.C.I} = \begin{cases} \left(\dfrac{m_{ij}}{\text{MIN}(m_i, m_j)} - \dfrac{\text{MAX}(m_i, m_j)}{M} \right) & \text{when } m_{ij} \geq 4. \\ 0 \quad \text{when } m_{ij} < 4 \end{cases}$$

where M = number of words in vocabulary = 1000
m_i, m_j = number of words associated with word i,j respectively
m_{ij} = number of words associated with both i and j.

FREQ. R	REAL ABSTRACTS		RANDOM ABSTRACTS	
	NO. PAIRS $P_1(R)$	CUMULATIVE TOTAL $\Sigma P_1(R)$	NO. PAIRS $P_2(R)$	CUMULATIVE TOTAL $\Sigma P_2(R)$
0	443,121	499,500	442,874	499,500
1	38,778	56,379	42,183	56,626
2	9,325	17,601	8,370	14,443
3	3,461	8,276	2,884	6,073
4	1,781	4,815	1,276	3,189
5	942	3,034	649	1,913
6	609	2,092	389	1,264
7	336	1,483	239	875
8	271	1,147	153	636
9	206	876	127	483
10	138	670	83	356
11	100	532	49	273
12	82	432	53	224
13	57	350	35	171
14	51	293	26	136
15	46	242	21	110
16	31	196	14	89
17	15	165	11	75
18	21	150	9	64
19	16	129	6	55
20	15	113	13	49
21	16	98	3	36
22	8	82	3	33
23	15	74	6	30
24	9	59	6	24
25	9	50	0	18
26	5	41	3	18
27	0	36	2	15
28	7	36	4	13
29	3	29	0	9
30	3	26	2	9
31	2	23	1	7
32	3	21	1	6
33	2	18	0	5
35	2	16	2	5
36	1	14	0	3
38	1	13	1	3
39	3	12	1	2
41	2	9	0	1
42	1	7	1	1
43	2	6	0	0
46	1	4	0	0
47	1	3	0	0
48	1	2	0	0
119	1	1	0	0

Fig. III.8.1 Distribution of word pair frequencies for real and random abstracts.

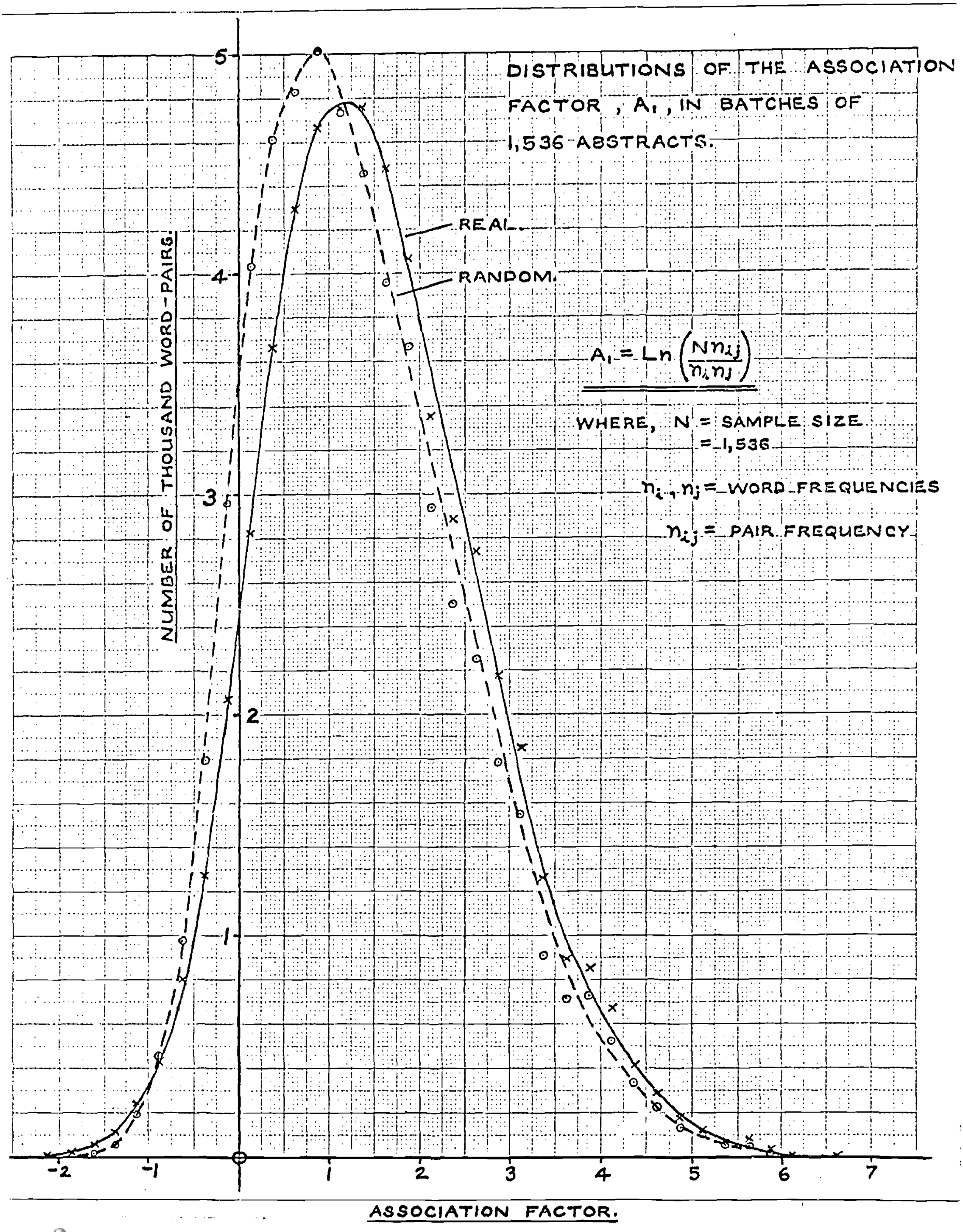

DISTRIBUTIONS OF THE ASSOCIATION FACTOR, A_1, IN BATCHES OF 1,536 ABSTRACTS.
REAL
RANDOM
$A_1 = \mathrm{Ln}\left(\frac{N n_{ij}}{n_i n_j}\right)$
WHERE, N = SAMPLE SIZE = 1,536
n_i, n_j = WORD FREQUENCIES
n_{ij} = PAIR FREQUENCY
NUMBER OF THOUSAND WORD-PAIRS.
ASSOCIATION FACTOR.
-2
-1
0
1
2
3
4
5
6
7

FIG. III.8.2

Note that the similarity coefficient is defined as being zero for pairs of words having fewer than four common associates. The matrix of these coefficients is stored on magnetic tape in the same way as the matrix of association factors.

III.10 Word connection matrices, G3 and G4

Two binary word connection matrices were used for producing word clusters for certain retrieval processes, in others they were used directly. One of these, G3, was produced, as a hybrid from both matrices A.F.5 and S.C.1. A complete print out of G3 appears in Appendix A5. The other, G4, was derived from S.C.1 alone (and, therefore, is just another word similarity matrix).

We thought it a good idea to produce one matrix as a hybrid so that any useful word connections brought out in the association matrix that may not be reflected in the similarity matrix might be retained. The general procedure used to form G3 was to threshold the A.F.5 and S.C.1 matrices and then to combine the resulting binary matrices to produce a new binary matrix containing a '1' where either one or both contain '1'. When thresholding, the simple process of setting a fixed threshold value for an entire matrix proved unsatisfactory. This way binary matrices were produced with an excessive number of 1's in same rows. These occurred in the case of words used very frequently in the abstracts, such as CIRCUIT, FREQUENCY, GEOPHYSICS, IONOSPHERE. The method used consisted in setting a nominal threshold value and, whenever a row was encountered having an excessive number of values above this threshold, raising the threshold <u>for that row</u> sufficiently to limit the number of 1's per row in the binary matrix produced to a specified maximum. The successive steps in producing G3 were:

1. Produce a binary matrix from S.C.1 using a nominal threshold value of 29/64, and limiting the number of 1's per row to 10.

2. Produce a binary matrix from A.F.5 using a nominal threshold value of 19/64, and limiting the number of 1's per row to 5.

3. Combine these binary matrices using a logical OR operation on pairs of corresponding elements, producing another binary matrix, X.

4. Transpose X.

5. Combine X and the transpose using a logical AND operation on pairs of corresponding elements.

6. Finally, produce G3 from the matrix produced in step 5 by giving all diagonal elements the value '1'.

Steps 4 and 5 are included to make G3 symmetric, the thresholding process in steps 1 and 2 destroying the symmetry of the original matrices. This thresholding gives higher priority to matrix S.C.1 than to A.F.5, only the highest association factors being permitted to influence G3.

After using G3 for some time we decided, on the basis of results produced, that it contained rather too many non-zero elements and that another matrix, G4, should be formed with fewer. We decided, partly on the basis of some evidence that most word pairs with high S.C.1 also have high A.F.5, to base this only on matrix S.C.1. We wanted to produce a matrix with a more uniform distribution of 1's in its rows than G3. For this reason the thresholding procedure was modified again. The steps involved in producing G4 were:

1. Produce a binary matrix from S.C.1 by setting a threshold value for each row so as to obtain a number of 1's in each row as close as possible to 7.

2. Transpose the resulting binary matrix.

3. Combine the matrix and the transpose using a logical AND operation on pairs of corresponding elements.

4. Finally, produce G4 from the matrix produced in step 3 by giving all diagonal elements the value '1'.

Matrix G3 has 188 rows in which only the diagonal element is non-zero (i.e. there are 188 completely isolated words), a maximum number of non-zero elements per row of 15 and a total of 3,582 non-zero elements. G4 has 112 rows in which only the diagonal element is non-zero, a maximum of 7 non-zero elements per row and a total of 3,074 non-zero elements.

IV Generation of Word Clusters

IV.1 Graph representation of word connections

The methods we use for word clustering assume as a starting point a set of binary word connections, i.e. connections which simply do or do not exist between words, having no associated weights or strengths. We have thought of these connections as being represented by a square binary matrix having rows and columns corresponding to the words and element values 0 or 1 to indicate the absence or presence of connections.

When considering problems of clustering we have found it useful to imagine this sort of data being represented by a graph. For this purpose we define a graph as a set of points and a set of lines connecting some or all pairs of points. Consider a graph having a unique point representing each word and lines representing a given set of word connections. The problem of finding clusters of words amounts to locating subsets of points in the graph which are fairly densely interconnected relative to the graph as a whole.

IV.2 Properties of clusters

The sort of properties of clusters in which we are interested is their number, size and degree of overlap. Having no means of knowing what values of these parameters are desirable for retrieval purposes, we decided to produce cluster sets exhibiting different characteristics and to test their retrieval capacities experimentally.

To some extent the properties of clusters derivable from a given set of word connection data are determined by the data. However, they are also influenced by the choice of definition of cluster. For example, a particular definition might not permit common membership of clusters, thereby restraining all clusters to be mutually exclusive. We have therefore tried three quite different methods of producing clusters, with three different implicit cluster definitions.

IV.3 Completeness of clusters

Each word in a completely connected cluster is connected to all other words in the cluster. That part of a graph corresponding to words and connections of a completely connected cluster forms a complete subgraph of the containing graph. If it is contained in no larger complete subgraph it is a maximally complete subgraph.*

In the case of word clusters based upon our statistical associations completely connected clusters containing more than three or four words are unlikely to emerge from the raw statistical data. Unless the connections specified by the raw data are supplemented in some way by further processing prior to cluster detection it is therefore sensible to adopt a definition of cluster which does not demand complete connectivity.

IV.4 Fragmentation of a graph by random removal of lines

The basic idea here is to break down the graph of connections into independent fragments by progressively removing lines from the graph, the

---

*Sometimes called a clique.

selection of lines to be removed being made with the aid of an algorithm for generating pseudo-random numbers. A set of fragments is independent in this sense if the points and lines in each are such that no connection exists between points in different fragments. An assumption underlying the method t. be described is that when a graph is broken down in this way connectivity is more likely to be preserved in those parts of the graph corresponding to clusters being sought (these parts of the graph having a relatively high density of connections) than elsewhere.

The graph defined by connection matrix G3 (section III.10) was examined to find its independent fragments. Apart from isolated words, there were two, one containing only four words, the other containing the rest. Independent fragments containing ten words or fewer were removed from the graph and 10% of the remaining connections were selected randomly and eliminated. The resulting graph was examined and again independent fragments containing ten words or fewer were removed. The processes of eliminating a random selection of connections and then removing from the graph any small independent fragments were repeated until the entire graph had been reduced to a number of such small fragments. Fragments obtained by this process containing from four to ten words each were stored. The complete procedure was repeated 21 times, each time starting with the graph of G3 and eliminating different random selections of connections, until 1,000 fragments were accumulated in the size range four to ten words.

At this stage we were likely to have obtained a number of fragments approximating to each cluster of words in the original graph. The processing that followed was intended to reconstitute the clusters by first comparing the 1,000 fragments accumulated, identifying sets of similar or overlapping fragments and merging them.

A few words featured rather insignificantly, each appearing in only one or two of the 1,000 fragments. These words were discarded. A similarity matrix was produced showing similarities between pairs of fragments. The measure of similarity used was the number of words that two fragments had in common divided by the number of words in the smaller. Different thresholds were set to produce binary similarity matrices which were processed to find independent sets of fragments, i.e. sets of similar fragments having no similarity connections between members of different sets. A similarity threshold of 19/32 was selected which maximized the number of independent sets of fragments, producing 148. Each set of fragments was merged to form a cluster by including in the cluster any word appearing in at least half of the fragments in the set. The word clusters thus produced were not mutually exclusive since the sets of fragments from which they were produced, though independent with a similarity threshold of 19/32, would not all have been independent had a lower threshold value been used.

446 distinct words appear in the 148 clusters formed. These clusters together with the 554 other words (used singly) constituted cluster set C4 used in retrieval run 19.

IV.5 Meetham's method of clustering, based upon removal of lines not included in increasing numbers of triangles

Here a basic assumption is that, for indexing and retrieval purposes, word groups should be mutually exclusive and that words which are associated with two otherwise weakly related or non-related word groups constitute an undesirable ambiguity.

Word connection matrix G3 was the starting point for this clustering experiment and, with the exception of 188 completely isolated words, it

consists largely of a set of words from each of which paths of connections can be traced to every other word in the set. The aim was to break down the graph of word connections into mutually exclusive components, the words in which satisfy intuitive notions of word clusters as regards their size and semantic range. The following procedure, used to obtain this break-down, is described in [10], while [8] and [11] are also of interest in this connection.

All lines (connections) were removed from G3 which were not included in at least one triangle (of connections). This split the graph into exclusive components containing 413, 298, 134, 81, 18, 12, 9, 7, 6, 5, 5, 3 and 3 lines respectively. Lines were removed from the four largest components if they were not included in at least two triangles. This yielded 15 components containing 3 to 18 lines and 7 larger ones, the largest having 167 lines. The process of removing lines unless they were included in at least two triangles was applied iteratively to the 7 larger components until no further changes occurred. (An interesting theorem about the convergence of this iterative process is considered in [11]). This yielded 14 components having 6 to 18 lines and 14 having 20 to 75 lines.

So far G3 had been split into 53 small but strongly connected components. It was still desirable somehow to partition the 14 larger ones. Inspection showed that the removal of one or a small number of points from each conveniently partitioned them further: the points removed corresponded to the ambiguous words mentioned earlier. It was then necessary to decide whether to leave those removed as further isolated points, whether to link them up in some way or whether to include them in any of the word groups.

The resulting set of word clusters (set C2 used in retrieval run 16) included 491 isolated words and 122 clusters each containing between 2 and 15 words.

IV.6 Cluster emphasis*

It is assumed that, in the graph of word connections, the subgraphs representing clusters being sought are unlikely always to be complete. An iterative procedure is used to process the original graph in such a way that clusters should be emphasised, connectivity within them being increased and that between them reduced, to a point at which it is possible to detect them as maximally complete subgraphs.+

Consider a binary connection matrix (section III.10), C, with 1,000 rows and columns corresponding to the word stems being clustered, and element values of 1 and 0 representing the presence and absence of connections respectively. Consider also the corresponding graph, G. In most clusters there will be pairs of words which are not directly connected. Some such pairs are likely to be connected by paths of length two (i.e. via a third word). The hypothesis underlying the technique described is that, considering all pairs

---

*An account of this part of the work was given at the 1968 IFIP Congress at Edinburgh and is published in the Proceedings [9].

+We used Wolfberg's algorithm for finding all the maximally complete subgraphs of a given graph. A reference to this, [22], appears in the bibliography.

of words not directly connected in G, the proportion of such pairs connected by paths of length two is likely to be higher for the set of pairs occurring within clusters than for other pairs.

The first step of the emphasis procedure is to form a new non-binary connection matrix the same size as C. A direct connection in G between a given pair of words contributes μ towards the value of the corresponding element in the new matrix, and <u>each</u> path of length two between the words contributes λ. The new matrix is computed from C as $(\mu C+\lambda C^2)$, integer values of μ and λ being specified when the program is run. A binary form, C', of the new matrix is produced by thresholding its values. The relative contributions of direct paths and paths of length two in G to the connections in the graph G', corresponding to C', depend upon the values chosen for μ and λ. Suppose that the threshold value is chosen to make the resulting number of connections in G' as close as possible to the number in G. If the above hypothesis is correct the density of connections within clusters will be higher in G' than in G and, of necessity, the density of connections elsewhere will be lower in G' than in G. The desired emphasis of clusters will thus be achieved.

Using G3 and G4 (section III.10) as initial connection matrices, this process was applied iteratively with various values of μ and λ and trying different methods of thresholding, including that outlined.

The process seemed fairly convergent in the cases tried. Two sets of maximally complete subgraphs (MCS's) were used as descriptors. One set, produced from G3, contains 1,178 highly overlapping clusters. This includes 188 isolated words. These were produced by a single application of the emphasis process with $\mu = 2$, $\lambda = 1$ and a threshold value of 2. This is the cluster set, C3, referred to in section VI.2.2 and used in retrieval run 14. A sample showing the very high degree of overlap of clusters in this set appears in Appendix A6. The set was selected because of this characteristic. The size distribution of these MCS's and of those of G3 appears in fig. IV.6.1. The size of the largest MCS is increased from 6 to 12 and the number of MCS's of 5 or more words is increased drastically by the emphasis process. However, the number of connections in the graph has increased approximately five-fold.

EMPHASIS	NO. OF NON-ZERO ELMTS.	NUMBER OF MCS's OF SIZE												TOT. NO. MCS's
		1	2	3	4	5	6	7	8	9	10	11	12	
BEFORE	3,582	188	786	284	88	18	1							1,365
AFTER	17,324	188	473	32	140	99	77	64	57	30	13	3	2	1,178

Fig. IV.6.1 Size distribution of MCS's before and after emphasis.

The second set, cluster set C5 used in retrieval run 20, was produced after five iterations of the process, starting with connection matrix G4, with $\mu = \lambda = 1$ and thresholding to maintain a number of 1's in the matrices as close as possible to the 3,074 in G4. Fig. IV.6.2 shows the distribution of sizes of the MCS's of the original graph and of those obtained after the iterations.

ITERATION	NO. OF NON-ZERO ELMTS.	NUMBER OF MCS's OF SIZ.																	TOT. NO. MCS's
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
0	3074	112	892	181	41	1	1												1228
1	2899	525	85	90	49	38	17	7	1										812
2	2981	525	79	79	24	19	18	8	10	5	2	2							771
3	2963	593	0	80	24	13	8	8	3	4	2	0	3	0	2	1			741
4	3095	593	0	76	19	13	8	5	2	6	1	0	1	0	0	0	2	1	727
5	3111	593	0	76	19	12	7	6	2	4	1	1	1	0	0	0	2	1	725

Fig. IV.6.2 Distribution of maximally complete subgraphs produced by five iterations.

The convergence evident from fig. IV.6.2 is typical. Two typical clusters obtained in the final iteration are shown in fig. IV.6.3, the connections being those in G4 before any emphasis.

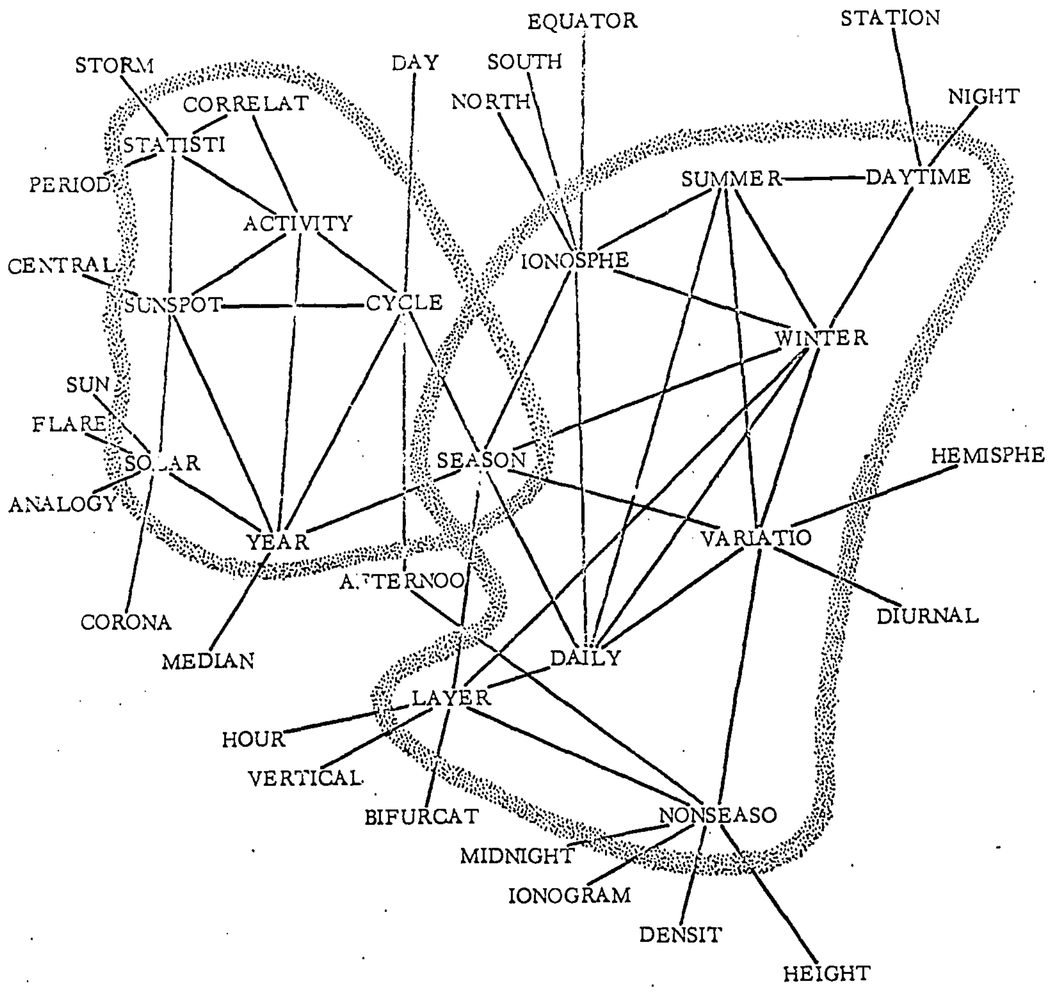


Fig. IV.6.3 Two clusters obtained from G4 by five iterations of the emphasis process.

PART 3 : Retrieval Experiments

V. General Considerations

V.1 Brief description of KDF9

KDF9 is a medium size, general purpose machine having a 48-bit word length. The main fast store, or core store, can be shared by up to four programs running at fixed priority levels determined by the operator. The core store has fetch and store times of six microseconds. The times taken for the fixed point operations of addition and multiplication, exclusive of fetch and store, are 1 and 14 microseconds respectively.

Special features of the machine are its two 16-word nesting stores, or stacks; one used to store the operands and results of any functions performed, the other used to hold the link instructions for nested subroutines.

The NPL machine has the following configuration:

32,000 word core store,
6 magnetic tape units,
4,000,000 word disc backing store,
Line printer, 120 character line, 600 lines/min.,
Paper tape reader, 1,000 characters/sec.,
Paper tape punch, 110 characters/sec.,
Card reader, 600 cards/min.,
Typewriter control terminal.

V.2 Search requests

When published, the abstracts used in these experiments appeared under five broad subject headings. The distribution was unbalanced even under these headings. We therefore had no knowledge of the subject profile of the collection. In collecting a set of search requests we naturally wished to achieve a reasonable correspondence between the distribution of their topics and that of the subjects in our document collection. We were thus faced with the problem of trying to match this unknown subject profile. We decided to do this by basing the requests upon a set of abstracts arbitrarily selected from the collection.

It was thought that a set of 100 requests would provide a reasonable spread of subjects. This number was obtained from 21 people, the number per person varying from two to seven. Each person was sent a set of source abstracts each of which he was told to consider simply as specifying the subject area of one search request. He was told that it was immaterial whether or not a source abstract was a relevant answer to the request based upon it.

The requestors were also informed that the systems under test were based entirely upon key-words and that all numerical data and mathematical symbols had been omitted when key-punching the abstracts. In accordance with this they were asked to exclude from their requests numerical data, symbols other than those in the Roman alphabet, and abbreviations.

A complete listing of the requests appears in Appendix B1. It is evident that the exclusion of numerical data has in some requests led to the use of words such as high and low where, for example, a frequency range or

DATE: 27/07/68 RUN 11 / REQ 20

W.C.BAIN RSRS REQUEST 20 SOURCE ABSTRACT 527

REQUEST: (011/026/009/004)

OBSERVATIONS OF RAPID FLUCTUATIONS IN THE EARTHS MAGNETIC FIELD AND THEIR RELATION TO THE PROPAGATION OF HYDROMAGNETIC W
AVES IN THE EXOSPHERE.

01141	RELATIONS BETWEEN THE ELECTRIC AND MAGNETIC FIELDS OF VERY LONG PERIOD INDUCED IN A MEDIUM OF VARIABLE CONDUCTIVITY. A MODEL OF GROUND CONDUCTIVITY DISTRIBUTION IS DERIVED WHICY CLOSELY REPRESENTS THE CONDITIONS INDICATED BY OBSERVATIONS OF GEOMAGNETIC AND GEOELECTRIC PULSATIONS.	/REQ./RUN./ ABST. /REL./ / 20 / 11 / 01141 / /
04543	ELECTROMAGNETIC INDUCTION IN A TWO LAYER EARTH. THE THEORY DEVELOPED BY IS EXTENDED TO DETERMINE THE INDUCED FIELD ON THE SURFACE OF A MEDIUM WHOSE CONDUCTIVITY IS A FUNCTION OF DEPTH. THE RESULTS ARE APPLIED TO EVALUATE THE FIELD INDUCED WHEN AN OSCILLATING MAGNETIC DIPOLE IS PLACED ON THE SURFACE OF A FLAT TWO LAYER EARTH.	/REQ./RUN./ ABST. /REL./ / 20 / 11 / 04543 / /
05417	INFLUENCE OF MAGNETIC FIELD ON CONVECTIVE INSTABILITY IN THE ATMOSPHERES OF STARS AND IN THE IONOSPHERE OF THE EARTH. THE CASE OF A MEDIUM WITH ANISOTROPIC ELECTRICAL AND THERMAL CONDUCTIVITIES IS CONSIDERED. IN A PLASMA WHICH DOES NOT CONTAIN A LARGE NUMBER OF NEUTRAL PARTICLES THE CALCULATED INFLUENCE OF THE MAGNETIC FIELD ON CONVECTION IS ESSENTIALLY DIFFERENT FROM THAT PREDICTED BY THE MAGNETOHYDRODYNAMIC APPROXIMATION IN THE IONOSPHERE THE INFLUENCE OF THE MAGNETIC FIELD IS SMALL OWING TO THE PRESENCE OF A LARGE NUMBER OF MOLECULES.	/REG./RUN./ ABST. /REL./ / 20 / 11 / 05417 / /
08400	MAGNETIC FIELD OF THE SUN. THE PRINCIPLES OF OPERATION OF BABCOCKS ZEEMAN EFFECT SOLAR MAGNETOGRAPH ARE DESCRIBED AND SOME EXPERIMENTAL RESULTS OBTAINED ARE SHOWN IN THE FORM OF MAGNETOGRAMS. THE GENERAL MAGNETIC FIELD STRENGTH IS OERSTED OF POLARITY OPPOSITE TO THAT OF THE EARTH. LARGE LOCAL BIPOLAR AND UNIPOLAR FIELDS HAVE BEEN OBSERVED AND ARE SHOWN IN SEVERL OF THE MAGNETOGRAMS. THE FIELD DISTRIBUTION VARIES FROM DAY TO DAY AND FLUCTUATIONS OF THE ORDER OF OERSTED ARE BELIEVED TO OCCUR IN LOCAL FIELDS WITH PERIODS OF ABOUT MIN. THE UNIPOLAR FIELDS ARE PROBABLY THE SOURCES OF CORPUSCULAR EMISSION WHICH PRODUCE GEOMAGNETIC STORMS WITH A DAY RECURRENCE PERIOD.	/REG./RUN./ ABST. /REL./ / 20 / 11 / 08400 / /

Fig. V.3.1. Sample of retrieved material as output from computer

attenuation factor might otherwise have been specified more precisely. However, this sort of thing is noticeable in relatively few cases.

Some time after eliciting the requests we wished to try out a retrieval strategy (see section VI.2.6) in which a specified word or words in a request were regarded as obligatory in that only abstracts containing such words would be retrieved. We asked the requestors to consider their requests and to indicate by underlining any word in them for which they were fairly confident that there was no suitable substitute.

Only 93 of the original set of 100 requests were used. However, the other seven are included in Appendix B1, together with the reasons for their cancellation, in Appendix B3.

V.3 Form of output

For each request, the output from each retrieval run includes the full title and abstract of all items retrieved in answer to the request that have not been so retrieved in any previous run. Thus, the first time a particular abstract is retrieved in response to a given request it gets printed and sent to the author of the request for relevance assessment. If the same abstract is retrieved in a later run in response to the same request it is not printed out. This minimizes the burden placed upon the requestors to assess relevance. It means, however, that the relevance assessments must be fed back to the computer for reference in later runs (see section V.7 on the harvest file).

A sample showing the output format appears in Figure V.3.1. The request is printed, followed by the retrieved abstracts. The first line of each abstract is actually the document title. The serial number of each abstract appears on the left, and to the right of each is repeated request number/run number/serial number, with a space for the requestor to insert a number indicating his relevance judgement. The output continues on as many pages as are required for the retrieved abstracts. In runs utilizing the underlining of request words the underlining is indicated in the print out by a string of X's beneath the word or words in question.

The ultimate page of the output for each request is a summary sheet which simply lists the serial numbers of all the abstracts printed out for the request with spaces for the insertion of relevance assessments. These summary sheets were not sent with the rest of the output to the requestors. When the relevance assessments for a run were available the information was copied on to the summary sheets. It was found useful to have the information in this compact form when preparing paper tapes to feed it back to the computer for storage in the harvest file (appendix B6).

V.4 Batch processing

The indexing and retrieval operations involve processing files, many of which are held on magnetic tape and some of which are rather large. The abstract/word matrix is an example of a large file, consisting of 11,571 lists of word numbers, one corresponding to each abstract indicating the words it contains. These are written on to magnetic tape in abstract serial number order. They occupy about 20% of a reel of tape and take nearly one minute to read or write. When such a file is being processed the time occupied by magnetic tape movements generally far exceeds the actual computing time incurred. This means that for much of the processing involved in

indexing and retrieval only a marginal increase in total computer time is required to process a batch of 100 requests instead of a single one. Consider, for example, the process involved in comparing the sets of key-words occurring in the abstracts with those occurring in the requests (the main part of the basic key-word strategy described in section VI.1). The 100 key-word sets in the requests occupy little space and can easily be held together in the core store. When this is done it requires only a single scan of the much larger file indicating the set of words in each abstract in order to compare all requests with all abstracts. The same amount of tape-scanning would be required to process a single request.

Another advantage of batch processing the requests is seen in the final stage of the retrieval process where it is necessary to assemble the particular selection of abstract texts to be printed out for each request. The complete file of texts fills a 2,300 ft reel of magnetic tape. It takes about four minutes to scan this file and hence that is approximately the time that would be required to assemble the abstracts to be output for a single request. With batch processing we begin by reading the entire file of abstract texts from tape and dumping them on to the magnetic disc. This operation takes about 12 minutes and would therefore be pointless when processing a single request. Once on the disc, the abstracts can be accessed randomly in very little time. The requests are then considered in sequence, the abstracts required for each are read into the machine, arranged in the required format and printed out.

V.5 Possible Variables

Various methods have been suggested for improving key-word stem performance independent of word-associations, not all of which we have wished to try.

The principal are:-

(1) Weighting individual request words, by reference to the requester.

(2) Weighting individual document words by counting their repetitions.

(3) Using functions such as the 'cosine' to measure the degree of match.

(4) 'Underlining', that is, insisting that one or more particular request words be present in each document retrieved.

To examine (1) thoroughly by varying the weights would involve much more work for the requesters than we were prepared to ask them to do. However, a small experiment was tried in getting the machine to assign weights, namely our Run 22 (AWKWS).

(2) seems of doubtful value with <u>short</u> abstracts where a chance repetition is more likely to introduce noise.

(3) takes account of the length of the document, so that e.g. a document only partly on the subject of the request scores lower than one wholly devoted to it.

These methods introduce new variables which we have not been able to try in addition to our own work on association techniques.

(4) though not appropriate to every request, we have explored, and we discuss it in section VI.2.6.

V.6 Quantity of output and cut-off

Consider these totals from Appendix B7:-

Run 13. Coordination = Number of Key-word stems in Common.

Request 0.	Coordination	Number of Documents	Output (Cumulative)
(6 keywords)	5	14	14
	4	87	101
	3	463	564
	2	1388	1952
	1	3639	5591
	0	5980	11571
Request 2.	4	2	2
	3	26	28
(5 keywords)	2	203	231
	1	1803	2034
	0	9537	11571
Request 3.	3	14	14
	2	228	242
(6 keywords)	1	2426	2668
	0	8903	11571

Fig. V.6.1

That is, 14 documents have 5 out of the 6 keywords of Request 0, 87 have 4, 101 have 5 or 4 and so on. (It happens that none have all 6). Since individual documents at any one coordination level are not distinguished, they must be output together. Retrieval consists in 'creaming off' successive 'strata' from the library, which has been arranged as it were in a 'pyramid'.

We shall use

stratum = non-empty set of documents at the same coordination

output (when used as a noun) = total (number of) documents, relevant or irrelevant, from the top coordination down to a given point inclusive, for one or more requests, and consisting of a number of complete strata, or zero.

As we have described in section I.14, the customer specifies a cut-off value, which we denote by K. He asks for as near as possible K documents, with, let us say, fewer rather than more in the case of a tie. Now 0 and 2K are equidistant from K. If the top stratum has 2K or more documents, the strategy is not sufficiently sensitive for the customer's purpose and gives no output. Thus in any case the output lies in the range 0 to (2K-1) inclusive.

Consider Fig. V.6.1 again, and set K = 20.

The outputs for requests 0,2,3,4,5,7,9, ..., 99 are 14, 28, 14, 9, 7, 12, 4, ..., 33 whose average is only 14.20, considerably less than 20.

Similarly, if we ask for 50 documents, we get an average of 35 per request.

We define

K' = effective cut-off = average number of documents per request output for a particular run and a particular request set.

In short, we ask for K documents per request and get K' on average.

By inspection we find that (1) for any given run and our 93 requests, $\frac{K'}{K}$ is approximately constant,

Approximate value of $\frac{K'}{K}$

	Strategy 13	Strategy 14
K = 20	.710	.919
30	.739	.931
40	.725	.918
50	.697	.926
100	.650	.959
250	.827	.971
500	.732	.956

Fig. V.6.2

and that (2) this constant is not necessarily unity.

Take an imaginary single request, with possible outputs of (say) 0, 3, 8, 15, ..., documents.

K	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...
K'	0	0	3	3	3	3	8	8	8	8	8	8	15	15	15	15	...

Fig. V.6.3

$\frac{K'}{K}$ oscillates about the value 1. If we have several requests, we might expect these oscillations to cancel to some extent, since the points at which the average K' changes become more frequent. This helps to explain (1) but (2) is surprising.

In Appendix B14 we show that (2) is a second-order effect which remains appreciable under the following sets of conditions, which, in our context, are roughly equivalent to each other:-

(i) For any given request the ratio between successive cumulative output totals is quite large, say 3:1 or more.

(ii) The outputs nearest to K for the different requests have a large spread.

(iii) The strategy employs few coordination levels.

In fact, if we can assume (1) we show there that

$$\frac{K'}{K} \simeq \frac{1}{1+v^2}$$

where $v = \frac{\text{standard deviation}}{\text{mean}}$

for the outputs nearest to K, so that v also is approximately constant when $\frac{K'}{K}$ is.

V.7 The sensitivity of a strategy to changes in the cut-off K

A number of ideas put forward in the last section relate particularly to systems such as ours where the documents fall into strata of several or many at a time. We say that a strategy is the more sensitive the more strata it has within a given output. For example, Run 13 (KWS) at K = 50, K' ≃ 35 has only 1.9 on average, and 2.0 at K = 71, K' ≃ 50. Run 14 (MCS01), a descriptor run, has 6.6 at K = 55, K' ≃ 50. Thus a given increase in K is more likely to yield new output for Run 14 than Run 13. This is important in an interactive use where documents are produced, let us say on a screen, stratum by stratum.

From V.6 it appears that the value of $\frac{K'}{K}$ if sufficiently constant, provides a useful guide to sensitivity, being nearer to unity the more sensitive the strategy. Closely related are the ratios of successive output totals for single requests, and the spread of the outputs required to approximate to K over a set of requests.

The effect of underlining (VI.2.6) is to remove part or all from each stratum. In general, the number of strata within a given output appears to decrease and the strategy becomes less sensitive. This is more marked when Run 13 is underlined (Run 25) than when Run 14 is (Run 21), as may be deduced from the tables of K and K' in Appendix B5.

V.8 The harvest file

This name was given to the file kept on magnetic tape and designed to store the results of every strategy as found, in as useful and flexible form as possible. It contained for each request complete lists of abstract numbers retrieved within the cut-off by each successive strategy, their coordination and **relevance,** and a flag to mask the first occurrence of an abstract for a particular request.

Standard programs recopied assessments known from previous strategies, and arranged that no abstract was sent for assessment twice in regard to a particular request. This kept the work of the assessors to a minimum, decreasing markedly as more strategies were run. It also allowed us to try out strategies related to combinations of earlier runs, even if the new abstracts were not sent for assessment, since we could calculate to what cut-off they were already fully assessed.

The harvest file was basic to all our evaluation programs. These included studying selected requests, smaller cut-offs, and the sets of abstracts retrieved by one strategy and not by another. A separate program calculated the minimum cut-off needed to retrieve each abstract by the particular strategy; we would recommend storing this also on the harvest file in any future experiment, as each new strategy is included. Fuller details are shown in Appendix B6 .

VI Indexing and Retrieval Strategies

VI.1 Description of strategies

A summary of this section will be found in Appendix B4 .

The main kinds of strategy used for indexing and retrieval are indicated below:

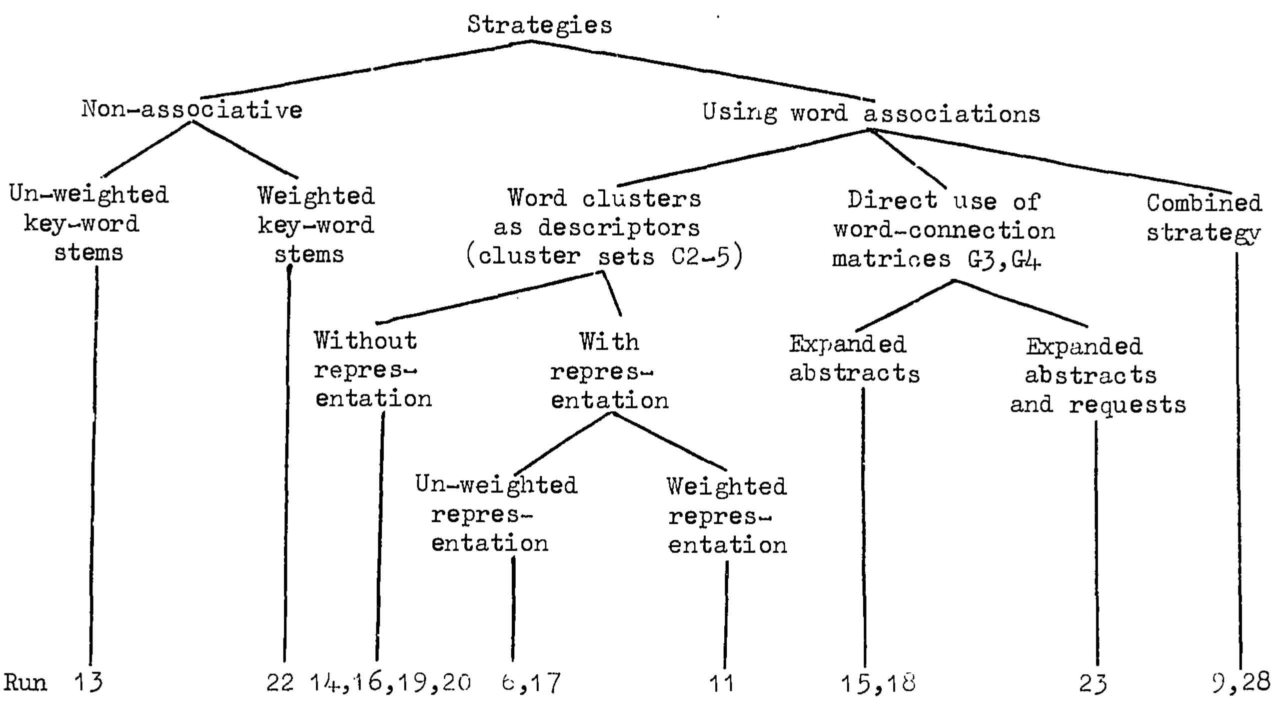


Fig. VI.1.1 Indexing and retrieval strategies employed

A complete description of the strategies follows. Most of the indexing and retrieval processes were implemented in terms of fairly basic matrix operations upon data arrays. Details of this matrix representation of the problem are not considered here, but are dealt with separately in Appendix B6, together with the computer programs used.

Determination of cut-off level

Our original idea was to aim at the same cut-off level, K, for each request in each run. Thus, in the first 13 runs (i.e. runs 0 to 12) the cut-off coordination value was always set so as to produce an output for each request as close as possible to 50 documents (i.e. K = 50).

As discussed in Section VII.3, we later decided that K', the average output per request, formed a better basis of comparison of different runs. Thus, in later runs, the average output was maintained as close as possible to 50 abstracts per request. Nine of the early runs were repeated in this way. In these cases only the repeat runs are quoted when presenting results.

The cut-off procedure thus involved carrying out the processing for a strategy to the point at which coordination values had been established for all documents in relation to each of the 93 requests being batch processed and forming tables similar to those in fig. V.6.1. Programs were then run which systematically tried different values of K, the number of documents aimed at per request, starting at one and increasing to find the value that produced an average output per request, K', closest to the desired value of 50. Having thus established the cut-off value to be set the necessary processing was continued to complete the particular strategy in hand.

The term standard runs is used to refer to the set of runs in which K' has been made as close as possible to 50, and the output from which has been fully assessed for relevance. The standard runs are 6, 9, 11 and 13 to 23 inclusive.

Basic key-word stem (KWS) strategy - run 13

The key-word stems in a request are compared with those occurring in each abstract. The number in common with a given abstract is noted and taken as its coordination value. The value of K is found giving a K' value closest to 50. The coordination value corresponding to this cut-off is determined and applied as a threshold to cream off the output for this request. All new material (i.e. abstracts in this output that have not previously been output for this request) is printed out for distribution to the requestor.

Simple use of word clusters as descriptors (no representation) - runs 14, 16, 19 and 20

The strategy here employs a set of word clusters, each cluster being used as a descriptor. The only difference between these runs is in the choice of word clusters, cluster sets C3, C2, C4 and C5 (see Section IV)being used in runs 14, 16, 19 and 20 respectively.

An abstract is indexed by assigning to it all clusters, from the particular set in use, with which it shares one or more key-word stems. Clusters are assigned in the same way to the requests according to their common key-word stems. From this point on the procedure is exactly analogous to the basic KWS strategy. The only difference is that the coordination value of an abstract in relation to a request is now the number of assigned clusters they have in common.

Use of clusters as descriptors with un-weighted representation - runs 6 and 17

Most of the sets of word clusters employed as descriptors are such that, although some words may appear in only one cluster, others occur in several. Thus, using the simple sort of strategy just described it is possible for two abstracts retrieved at the same descriptor coordination level to represent, by inclusion of the actual words or assignment of clusters containing them, different numbers of request words. We define the representation level of an abstract in relation to a given request as the number of request key-word stems actually occurring in the abstract or appearing in at least one cluster assigned to index it.

RADIO NOISE FROM PLANETS - F. Horner. (Nature, London, vol. 180, p. 1253; Dec. 7, 1957). A comparison is made of H.F. radio noise from Jupiter and Venus and from terrestrial lightning. The hypothesis that radio noise from the two planets is due to electrical discharges analogous to terrestrial lightning requires modification.

ABSTRACT A

AN INVESTIGATION OF WHISTLING ATMOSPHERICS - L.R.O. Storey. (Phil. Trans. A., vol. 246, pp. 113-141; July 9, 1953). A comprehensive report of an experimental and theoretical study of whistling atmospherics, at frequencies $<$15 kc is given. Whistlers may or may not be preceded by ordinary atmospherics, produced by lightning strokes at a distance of $\sim$2000 km. The diurnal and annual variations of the properties of both types were investigated. Explanatory theory of their origin advanced by Eckersley is developed. Measurements of the degree of dispersion indicate an electron density in the upper atmosphere considerably larger than expected. This result is explained on the assumption that electrons are falling in from outside, and this might account for the relation between the occurrence of whistlers and magnetic activity.

ABSTRACT B

Fig. VI.1.2

An example is provided by request no. 36 for abstracts on simultaneous observations of whistlers and lightning discharges. Considering cluster set C3 the key-words, which are underlined, occur as follows:-

SIMULTANEOUS	in clusters	712 and 713
OBSERVATION	" "	465, 466, 467, 468, 469, 595, 596, 597, 598, 599 and 600
WHISTLER	" "	326 and 1032
LIGHTNING	" "	110, 279, 546 and 547
DISCHARGE	" "	279, 280, 281 and 282

Thus, the request is formulated by 22 distinct descriptor clusters.

In relation to this request it is interesting to note that (referring to Fig. VI.1.2):

Abstract A - contains 2 key-words of the request, has descriptor coordination level 14, has representation level 3.

Abstract B - contains 2 key-words of the request, has descriptor coordination level 7, has representation level 4.

Thus, although Abstract A is indexed by twice as many request descriptors as Abstract B, it represents fewer of the key-words in the request.

This strategy attributes importance to the representation level. The complete set of abstracts is ordered, placing abstracts representing the largest number of request words at the top of the list, followed by those representing the next largest number, etc. Abstracts with the same representation level are ordered, as in the previous strategy, according to their descriptor coordination level. The process of creaming off as nearly as possible to the required number of abstracts is carried out as in the KWS strategy.

Use of clusters as descriptors with weighted representation - run 11

In the previous strategy an abstract, A, represents a request word, W, if it is indexed by at least one word cluster (drawn from the set being used as descriptors) containing W. It seems reasonable to regard W as being more strongly, or better, represented by abstracts indexed by several clusters containing W than by only one. Accordingly, the criterion for representation is now modified slightly. An abstract represents a request word if it is indexed by at least one third of the clusters containing the word. Apart from this the strategy remains unchanged.

Word-stem comparison with expai ed abstracts - runs 15 and 18

One possible advantage to be gained by em loying word clusters to reflect word associations is economy of storage space in the computer. Rather than store all the individual word-pair associations one simply keeps a record of the word content of each cluster. In the case of a cluster containing n words this means storing n items as against some number almost certainly in excess of this and, possibly, as great as n(n-1) if the individual associations are stored. There is also the hope, which has a great deal of intuitive appeal, **that the clusters one finds will represent in some way salient** themes or concepts relating to the subject matter of one's document collection.

However, these possible advantages could well be offset by the undeniable fact that sensible clustering procedures, even crude ones, are rather costly in terms of computer time.

We therefore thought it worthwhile testing procedures which do not involve clusters at all, but make direct reference to a word connection matrix. Two such matrices were used, matrix G3 in run 15 and matrix G4 in run 18.

Each abstract is processed by noting its key-word stems, referring to the connection matrix to find all the words connected to each of them, and adding these to the original words to form an expanded abstract. Key-words in the requests are then compared with those in the expanded abstracts, continuing as with the basic KWS strategy.

Word-stem comparison with expanded abstracts and requests - run 23

Using connection matrix G4 the same method of expansion is applied to both the abstracts and the requests. The procedure is then identical with the basic KWS strategy.

Combined strategy - run 9

This is intended to combine the advantages of the basic KWS strategy (run 13) and the strategy of run 14 which employs the set of word clusters, C3, and performs relatively well. These two runs are effectively repeated using a K' value of 500 instead of the usual 50. For each request a new list of abstract numbers is produced containing those appearing in both of the sets of approximately 500 thus produced. A new "coordination value" is assigned to each abstract listed obtained as the product of the values in runs 13 and 14. Finally a K' value of 50 is taken to determine the cut-off for each request resulting in an average output of 50 abstracts per request.

Combined strategy - run 28

This is a combination of the basic KWS strategy (run 13) and that used

in run 14. Coordination values are obtained as follows:

$$\text{Run 28 coord}^{n}\text{. value} = 2^{7} \times (\text{run 13 coord}^{n}\text{. value}) + (\text{run 14 coord}^{n}\text{. value}).$$

The effect is to retain the major division of the document collection produced in run 13. However, the document sets at each coordination level in run 13 are now ordered according to the coordination levels they obtained in run 14. The factor 2^{7} in the first term is greater than 99, the highest coordination value in run 14, thus avoiding any carry between the terms in the above expression.

Obligatory request words - runs 21, 24 and 25

Information regarding the word(s) in his request which, given the option, a requestor would make obligatory was used to remove from consideration all abstracts not containing the word(s). The strategies of runs 10, 11 and 13 were then repeated, in runs 21, 24 and 25 respectively, but instead of retrieving from the complete set of abstracts the appropriate reduced set was used for each request.

Automatic Weighting of Request Key-Word Stems - run 22

A set of programs was written to display the request words which had been involved in the retrieval of particular abstracts, whether directly (e.g. Run 0, i.e. the basic KWS strategy with K' = 35) or via some associated word (e.g. Run 11). It appeared, especially in the case of Run 0, that the request words often fell into two groups.

Consider, for example, Request 53.

ABSTRACT		INFORMAT	TRANSIST	AMPLIF	DESIGN	DRIFT	LOW
176	2		314	491	520		982
973	2		314	491	520		982
1504	3		314	491		692	982
1644	3		314	491	520	692	
1646	2		314	491	520		982
3430	2		314	491	520		982
3567	2		314	491	520		982
4498	2		314	491	520		982
5370	2		314	491	520		982
6202	2		314	491	520		982
7019	3			491	520	692	982
8030	3		314	491	520		982
8590	3		314	491	520		982
8694	2		314	491	520		982
8695	2		314	491	520		982
10505	3		314	491		692	982
10645	2		314	491	520	692	
10776	2	136	314	491	520		
11467	3		314	491	520	692	
11468	3		314	491		692	982

An indentation marks the abstracts later assessed as relevant (3 = relevant, 2 = irrelevant) and the numbers in the columns are simply the codes for the key-words at their head.

We notice that, if an abstract has 4 out of the 6 possible **key-words,** it is highly likely that 3 of them are Design Transistor and Amplifier, and also that the fourth is Low. The first 3 words define a background subject area, not in itself sufficiently relevant. On the other hand, the relevance is apparently much higher when DRIFT occurs which, in this request, is a highly selective word. We noticed a marked tendency for abstracts containing words appearing infrequently in the Run 0 output to be assessed relevant. It is tempting to apply this model generally and to suppose that these words are the more selective, and if weighted would lead to higher precision.

Lists were made mechanically for each request of request key-words occurring in 80% or more and in 90% or more of the output for Run 0, 80% seemed the better dividing line between the two classes of word.

The strategy of run 22 is similar to the basic KWS strategy in that the key-word stems of a request are compared with those of each abstract, no associations being considered. Words that occurred in less than 80% of the output for Run 0 contributed 2 to the "coordination value" instead of 1.

VI.2 Discussion of strategies

VI.2.1 Methods of using word associations

The associative strategies fall into two main groups corresponding to the adopted methods of employing information on word association and similarity measures, which are:

(i) use of word clusters, derived from the word associations, as descriptors for indexing and for comparing indexed abstracts with requests, and

(ii) direct use of the word association factors and similarity coefficients in the indexing and retrieval processes.

Using method (i) five different sets of clusters were tried. Though derived from the same basic set of word associations, they were produced by a variety of methods and they exhibit quite distinct characteristics. For example, the clusters in set C4 are mutually exclusive whereas some of the other sets contain considerable overlap of words. The number of clusters also varies considerably between the sets.

Method (ii) avoids the problems and effort involved in generating word clusters. It was used in three strategies to provide, by means of comparison, some indication of the benefits of clustering.

VI.2.2 Discussion of our use of word clusters

Clusters used as descriptors are assigned in the simplest possible ay to both documents and requests, one or more word stems in common between a document or request and a cluster being the sole requirement. Furthermore no explicit system of weighting is used, equal importance being attached to all assigned clusters. It would therefore appear that the assumption is being made that all the wo ds in a particular cluster are inter-substitutable in the given texts for retrieval purposes. This is true only in the case of the single set of mutually exclusive clusters. The position is not so simple in the case of clusters of words which overlap.

Consider the following situation in which A,B,C, etc., are words arranged in overlapping clusters, all the clusters containing A being

(A,B,C,D)
(A,B,E,F)
(A,B,G)
(A,B,D,E,H).

An abstract containing any of the above words would have one or more of these clusters assigned to it. Notice that if an abstract contained word A or B then all four clusters would be assigned. Word D, on the other hand, would only lead to the assignment of two of the clusters. Thus, the simple use in this way of overlapping clusters provides an in-built system of word weighting. Given the occurrence of word A in a search request, words appearing in abstracts are weighted as follows by virtue of the above clusters:

A	4
B	4
C	1
D	2
E	2
F	1
G	1
H	1.

As a practical example of this we have considered, in cluster set C3, all those containing the word stem STOR. There are 1,178 clusters in C3, having a high degree of overlap. The following list shows all the word stems appearing with STOR in at least one cluster and the respective weights they would receive for a request containing the stem STOR:

STOR	14	MEMOR	5
BIT	13	SPEED	5
DIGIT	11	DECIMAL	4
INFORMAT	11	MACHINE	4
DRUM	10	SYSTEM	4
ACCESS	9	LOGIC	2
COMPUT	7	CRYOTRON	1
READ	7	HEAD	1
TAPE	7	PRESET	1
READING	6		

A set of clusters with little overlap will produce correspondingly less refined word weightings.

VI.2.3 <u>Comparison with random superimposed coding</u>

It is interesting to compare our use of clusters of associated words with Mooers' system of Zatocoding or superimposed coding [16]. In this system key words used to index a document are represented by cutting notches in the edges of a card uniquely representing the document. The number of positions in which notches can be made is usually in the region of 30 to 50. In order to accommodate a key-word vocabulary probably far in excess of this number each key-word is represented not by one particular notch position, but by a unique set of notch positions. Usually about four positions are used for each word, and these are decided by using random number tables. The words in a search request are similarly transformed into a set of notch positions and documents are retrieved corresponding to cards having notches in the desired positions.

Three observations can be made:

(i) Since documents and requests are ultimately "described" by sets of notch positions, the total set of available notch positions functions as a set of descriptors,

(ii) The transformation of key-words into descriptors is non-unique in the sense that different <u>sets</u> of key-words might be transformed into the same set of descriptors, and

(iii) The transformation of key-words into descriptors is completely arbitrary.

This forms a very close parallel to our use of word clusters as descriptors, each notch corresponding to a word cluster.
The use of mutually exclusive clusters corresponds in Zatocoding to a situation in which each key-word is represented by a single notch position.
The two main differences between the use of overlapping clusters and Zatocoding are:

(i) Whereas all key-words are represented by the same number of notch positions using Zatocoding, in our system different words occur in different numbers of clusters, and

(ii) In our word cluster systems the relationship between key-words and descriptors is <u>not</u> arbitrary. On the contrary, it is carefully contrived so that associated words are clustered.

Our method of employing word clusters as descriptors may thus be thought of as a generalization of Zatocoding, or superimposed coding, in which the coding is associative or probabilistic rather than random.

VI.2.4 <u>Representation</u>

Consider a search request containing three key-word stems p_1, q_1, r_1. Let the set of all word stems appearing in at least one cluster containing p_1 be $\{p_1, p_2, p_3, p_4\}$, in those containing q_1 be $\{q_1, q_2\}$ and in those containing r_1 be $\{r_1, r_2, r_3\}$. The strategies employing un-weighted representation effectively formulate this request as

$$(p_1+p_2+p_3+p_4)\circ(q_1+q_2)\circ(r_1+r_2+r_3),$$

where '+' and '$\circ$' signify the logical operations of 'OR' and 'AND' respectively.

We are saying, for example, that an abstract containing $p_1 \circ q_2 \circ r$ [illegible] better than one containing $p_1 \cdot q_1 \circ q_2$, because the former represents all three request words whereas the latter represents only two. However we also imply, for example, that an abstract containing $p_1 \cdot p_3 \circ p_4 \cdot$ [illegible] $r_2 \circ r_3$ is <u>no</u> better than one containing $p_3 \circ q_2 \circ r_1$, as these represent the same number of request words.

The in-built word weightings produced by using overlapping clusters vanish when representation is introduced. The occurrence in an abstract of an actual request word contributes no more to the representation value of the abstract than the occurrence of another word clustered with a request word.

VI.2.5 <u>Direct use of word associations</u>

A word-connection matrix (Section III.10) is used to expand the request/abstract texts by addition of all connected words. Retrieval is then performed by word comparison, the coordination level being the number of common key-words. The words connected with a given one are regarded as possible substitutes for it for the purpose of retrieval. The first thing we tried, therefore, was expanding the abstract texts and comparing them with the (unexpanded) requests. The effect of doing this and then retrieving by key-word coordination is to accept connected words as substitutes (Runs 15 and 18).

At first sight it does not seem logical to expand the request texts in this way and then retrieve on the basis of key-word coordination. In this

case the maximum possible coordination value would equal the number of words in the expanded request and, in effect, the search would be for abstracts (or expanded ones) containing all words in the request and all words connected with them. However, in the case when the abstract texts are also expanded the effect is to produce a system of word weighting similar to that obtained using overlapping clusters. If an abstract contains a key-word in a request, then the expanded versions of both contain that word and all words connected with it, and they all contribute to the coordination value. A word in an abstract, not present in a request, may be thought of as having a weight in relation to a request word equal to the number of words with which they are commonly connected. This is their contribution to the coordination value. This method was tried in run 23.

VI.2.6 Obligatory request words

It seems generally to be sound practice for a search strategy to accept words associated with those in a request, or at least to attribute some weight to them. However, in discussing the matter with the requestors we found that occasionally a word was used for which the requestor could think of no useful substitute. It is quite likely in such a case that the word in question has associations with other words which would be misleading in the given context.

Words in all requests were considered by their authors and those whose presence in retrieved material was felt in this way to be obligatory wer underlined. Some strategies were then repeated with a slight modification so that abstracts not containing the stems of underlined words were not retrieved.

VII Evaluation Procedure

VII.1 Relevance

Each abstract was assessed as relevant or irrelevant by the subject expert who contributed the original request. The following code was used:-

3 Relevant, would be likely to contribute to a collection of documents answering the question.

2 Irrelevant, would not be likely to do so, not worth following up.

1 Unassessable owing to lack of numerical data or the like.

0 Not yet assessed.

All abstracts were keypunched omitting numerical or symbolic data, equations, formulae, or bibliographic references. This was because, with the keypunching available under contract for the first batch of abstracts, it was very difficult to check such data to a reasonable accuracy. Further, it was hoped that this omission would not greatly affect our examination of a system based essentially on key-words, and this has turned out to be so.

It was strictly for abstracts deficient for this reason that category '1' was made available. However, out of some 17,000 assessments, about 2,000 were '3' while there were no less than 500 '1's. Clearly some assessors were using it to indicate a lower degree of relevance and a letter was sent out asking for a firm decision wherever possible.
Finally 83 '1's remained, and in what follows these have been included in the '2's for simplicity's sake.

Note that the machine does not offer a relevance judgement. It may output an abstract for a certain cut-off K, but not at a lower K-value. The assessors on their part used subjective considerations, independently, for example, of the number of request words present.

VII.2 Precision

The customer specifies a cut-off value which we denote by K. He asks for as near as possible K documents, with, let us say, fewer rather than more in the case of a tie. Now 0 and 2K are equidistant from K. If the top stratum has 2K or more documents, the strategy is not sufficiently sensitive for the customer's purpose and gives no output. Thus in any case $0 \leq \text{output} \leq 2K-1$, where

(i) Output = Total number of Documents Retrieved on a Particular Occasion.

We shall use

(ii) Relevant Output = Number of Documents in Output Judged Relevant by User

and define (as is usual)

(iii) Precision $= \dfrac{\text{Relevant Output}}{\text{Output}}$.

If a customer increases the K-value the output will increase or not, according to the relative size of the next stratum. If it does, he may get more relevant documents, but usually he must expect them to be 'thinner on the ground'. For the machine can respond only by lowering the acceptable standard of matching, and this tends to lower precision as just defined.

To compare two strategies we take a set of requests, so that totals or averages may compensate for individual variations. Originally we assumed that identical K for both strategies would yield much the same total or average output, but this turned out not to be so (below section VII.3). Accordingly the K values were adjusted so that each strategy yielded 50 documents or nearly on average. Within one strategy K is still held constant.

We then calculate

(iv) Total Relevant Output over all Requests
(v) Precision for all Requests' Output Combined
(vi) Average Precision Ratio.

This is not, however, the whole story. A key-word or key-word stem (KWS) match is the simplest, and, involving no associations, remains the cheapest to set up and operate. A customer will want to know whether the extra cost of associations is worthwhile. Moreover, many documents will score highly in both types of run.

Accordingly, we are particularly interested in the numbers of new documents which could not be found by a basic KWS strategy in a reasonably sized output.

We calculate

(vii) Total New Relevant Output, i.e. not also retrieved in KWS at same or related K-value
(viii) Precision Ratio of New Output, taken separately
(ix) The number of requests which make a positive contribution to (i).

VII. 3 Operational and evaluation parameters, K and K'

Here a parameter is simply a variable which we can conveniently control in one or more strategies, in order to alter or compare their performance.

Examine the following figures:-

Request Set	Run	K	Relevant Abstracts	Irrelevant Abstracts	Output	K'	Precision
93	13(KWS)	50	799	2443	3242	35	25%
93	9(13T14)	50	972	3842	4814	52	20%
93	23(EARG4)	50	871	3441	4312	46	20%

These show for three different runs the response when the cut-off K is set at 50. Of the three, Run 9 produces most relevant documents. However,

the customers have had to scan half as much material again as in Run 13 in order to find them. Now, as we have said, the physical quantity of output is one of the chief factors in determining the cost and convenience of an operational system. Thus while cut-off K is the natural operating parameter, average output K' is the more appropriate evaluation parameter.

Reset the K-values till the K' are as near 50 as possible, rather than the K. Compare then

Request Set	Run	K	Relevant Abstracts	Irrelevant Abstracts	Output	K'	Precision
93	13(KWS)	71	991	3637	4628	50	21%
93	9(13T14)	49	966	3729	4695	50	21%
93	23(EARG4)	53	914	3793	4707	51	19%

Here Run 13 has most relevant documents. Also,

If we compare strategies at the same K' then precision is proportional to the number of relevant documents retrieved, and we need not quote both.

VII.4 Recall estimates

For a given request, let us define

(i) Perfect Recall = Number of Relevant Documents in Collection and, given in addition a Strategy and Cut-off,

(ii) $$\text{Recall Ratio} = \frac{\text{Relevant Output}}{\text{Perfect Recall}}$$

We have mentioned in Part I the difficulty of estimating perfect recall; let us write

(iii) Known Recall = Number of Relevant Documents retrieved by at least one of the 14 standard strategies

(iv) $$\text{Known Recall Ratio} = \frac{\text{Relevant Output}}{\text{Known Recall}}.$$

Largely manual searches of two types were made to test how meaningful these were.

Virtually Exhaustive Search, 3 requests, (Run 27)

The library was scanned as follows:-

1. Retain mechanically all documents having one or more words present in or associated with the request via the word-word association matrix G3, and print out those not previously assessed.

2. Mark documents having even the slightest connection with subject matter of request.

3. Send to requestors for thorough assessment.

Step 1 retained 5,000 documents on average, Step 2 about 200.

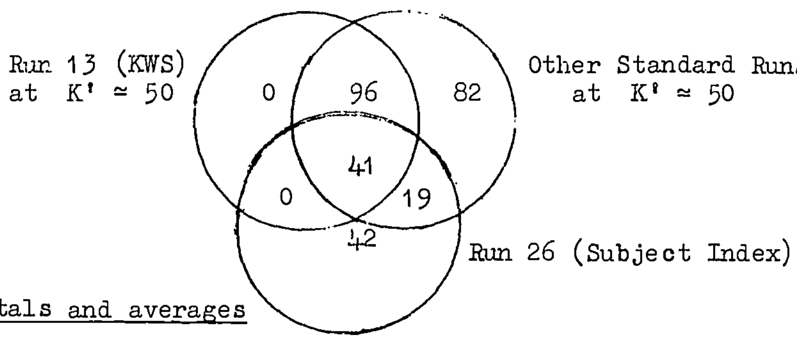
Relevant Documents found

	Run 13(KWS)	Other Standard Runs	Run 27 EXHSEA	Totals
Request 10	55	20	11	86
38	4	1	[illegible]	9
91	9	14	9	32
	68	35	24	127

If these figures are at all typical, then our combined runs have retrieved of the order of 80% of the relevant documents in the collection. Clearly the figures 11, 4, 9 in the fourth column are too small to have been estimated by a random sample much smaller than the whole library itself.

Subject Index Search, 12 requests, (Run 26)

Appendix B16 shows a page from the annual subject indexes. The headings vary little from year to year, while the entries are based on the abstract as a whole but are often rearrangements of words in the title. 12 of the requests were searched by their requestors under headings chosen by themselves. Since the entries did not give a very full picture the abstracts were found and printed out, and some 30% were then found to be irrelevant. Of the 102 relevant, 41 were known through Run 13 at $K' \simeq 50$, and a further 19 through the other standard runs. Thus 42 were quite new, against 238 already known, suggesting that the standard runs accounted for 80% or more of the total. The following diagram shows the numbers of relevant documents involved.



VII.5 Totals and averages

We have explained our reasons for using K or K' as our independent parameter, and numbers of documents, or precision for given output, as our chief evaluation measure.

For interest's sake we have added, in the runs tabulated in Appendix B8, three other measures which are worth considering. Overall Precision and Overall Known Recall refer to an imaginary request whose numbers of documents are the numbers for each request in the particular set added together. Thus Total Relevant and Overall Precision are what we have used already.

If for fixed K request i retrieves x_i relevant in y_i output out of z_i known relevant then

$$\text{Overall Precision} = \frac{\Sigma x_i}{\Sigma y_i} \quad \text{Overall Known Recall Ratio} = \frac{\Sigma x_i}{\Sigma z_i}$$

Average Precision $= \frac{1}{n} \Sigma \frac{x_i}{y_i}$ Average Known Recall Ratio $= \frac{1}{n} \Sigma \frac{x_i}{y_i}$

where n is the number of requests.

With the idea that a strategy yielding $x_i = y_i = 0$ is not very selective we have set precision equal to zero in such a case. (It happens that $z_i > 0$ for all our 93 requests). The smaller the value of K the greater the chance that $y_i = 0$ and consequently the average precisions actually increase for the first few K. The two recalls, and the two precisions after the initial rise in average precision, are not much different in size. There does, however, seem to be a systematic difference of a few per cent so that

$$\text{Overall Precision} < \text{Average Precision in general} \qquad \ldots \text{(VII.5.1)}$$

Now

$$\text{Overall Precision} = \frac{\Sigma \frac{x_i}{y_i} \cdot y_i}{\Sigma y_i}, \qquad \ldots \text{(VII.5.2)}$$

that is, it is an average precision in which precision ratio $\frac{x_i}{y_i}$ is given weight y_i. The outputs y_i are approximations to K but have a definite spread. Whether the output for an individual request is greater or less then K is likely to be independent of the request's particular output-precision curve. Hence we might expect that requests whose output went above K had a lower average precision than those that went below. Consequently, in (VII.5.2) the lower precisions would carry greater weight and this is what we have found in (VII.5.1). In confirmation, we may verify that the effect is much less marked in Runs 14 or 28 than in Runs 13 or 25, where the output spread is greater.

The same effect was noticed by Cleverdon [15] Volume II, pages 53-55 for certain of his measures, which were calculated by keeping key-word stem coordination constant and then totalling or averaging over a set of requests. It would not be surprising if, for a fixed coordination, large outputs correlated with small precision, and this is what Cleverdon's observation implies.

VIII Discussion of Performance Figures

VIII.1 General Comments

There is naturally no single answer to the question 'which is the best strategy?' Performances differ from request to request; it may not be worth going on to a second strategy, altering the cut-off, or using underlining. Broadly speaking we have in mind the person whose requests are batch processed along with many others, so that feedback is likely to be limited compared with an interactive system with an on-line terminal.

We shall use the following terms:-

Basic Strategy - Run 13

Standard Strategy - Runs 6, 9, 11, 13 - 23, fourteen in all.

A cut-off value of $K' \simeq 50$ is implied unless the contrary is stated. Assessments for runs 24, 25 and 28 are known only where these overlap the standard runs, and we write *25, etc. where the assessments are incomplete. However, Run 28 is almost fully assessed at $K' \simeq 50$ and we shall not go far wrong if we take the unassessed for this K' as irrelevant. We have therefore included Run 28 in most of our tables.

We asked for requests to be formulated in ordinary natural language. One or two contained words such as 'information' used in the non-technical sense, and which we thought might produce noise in the system. It appeared however that say 4 other key-words were sufficient to define a subject area so that the presence of such a key-word seemed to have little effect.

We were not in a position to work on improving or altering the wording of requests to see how this affected performance, other than by 'underlining' (VI.2.6). However, we picked out those requests which we felt were less well formulated (Appendix B3). It turned out that the precision ratios of these 17 were significantly lower in Run 13 than of the remaining 76, at the 5% level, but we did not feel justified in cancelling them also.

VIII.2 Relative Performance of Standard Strategies

Appendix B9 shows the numbers of relevant abstracts output by each of the standard strategies for all 93 requests for a selection of K'. It appears that of these 13 or 28 is the best run uniformly as K' varies. This is far from being the whole story, however. Some of the runs which do nearly as well as Run 13 do so because their output is little different. Since Run 13 does as well as any, let us assume that a requestor is likely to make it his first try.

Consider what happens if he goes on to try a second strategy. The machine is programmed to suppress all documents he has already received and presumably scanned. Appendix B9a shows the <u>new</u> documents output, that is, not also output in Run 13. Thus for example at $K' \simeq 50$, Run 9 produced 966 documents as against 991 documents for Run 13. Only 176 of these were new. Run 6 produced 753 but 238 of these were new. Again Run 21 produced 902 documents of which 311 were new and so is better than Run 6 on both counts. Consider the run numbers rank ordered by averaging their rank orders at $K' \simeq 20, 30, 40, 50$:-

93 Requests: Run Numbers in Averaged Rank Order

	(highest)														(lowest)
Total Relevant Output:	13	*28	9	22	23	19	21	20	11	14	18	15	16	17	6
New Relevant Output:		21	18	15 = 23	6	11 = 14	17 = 16	19	20	22	*28	9			

There is a weak anticorrelation between these two orders (omitting 13 in the first set), namely $\rho = -.37$, so that, roughly speaking, the nearer a run is in performance to Run 13 the fewer new documents it produces.

It is noticeable that some runs achieve a comparatively high total by doing very well for fewer requests. Thus Run 19 produced 218 new documents, but only 44 out of 93 requests contributed to this total. This coverage or number of productive requests is noted in brackets in the appendices where it is known. The coverages for one relevant document differ little,+ but those for new relevant documents are perhaps worth putting in average rank order:-

93 Requests: Run Numbers in Rank order

(highest) (lowest)

Coverage for New Documents: 6 23 11 18 15 21 14 16 = 20 17 19 9 *28 22

This seems to correlate to a fair extent with the numbers of new documents retrieved ($\rho = .80$).

The numbers of new documents are an important test of an association strategy's performance. The motive for going on to a second strategy is much stronger when Run 13 produces only a few documents. It does not necessarily follow that in such a case other runs are likely to produce even fewer. 34 requests, 37 per cent of our set, produced only 0 to 4 relevant documents in Run 13 at $K' \simeq 50$, (Appendix B10). Even if we omit the less well formulated requests, there are still 13 for which we need output of 250 or more if we are to get further relevant documents. New documents however are often retrieved for this set by association or descriptor runs at low K.

If we take the standard runs at $K' \simeq 50$ and then select the 34 requests, we get the figures in Appendix B11 and the following rank orders:-

34 Requests: Run Numbers in Rank Order

	(most)														(fewest)	
Total Relevant Output:	15	18	11	9	23	*28	21	20	14	6	17	16	19	*25	22	13
New Relevant Output:	21	15	18	23	11	9	14	*28	6	20	17	16	19	*25	22	
Coverage for New Documents:	18	11	23	6	15	21	9	14	17	16	20	*28	19	22		

- -

+Compare, however:-

Strategy	28	11 =	9	23	13	6 =	15 =	18 =	21 =	22	20	14 =	19	17	16
Numbers of Requests with 5 or more relevant documents	65	64	64	63	59	58	58	58	58	58	57	55	55	53	51

When we select requests at random we might expect the K' to remain roughly the same, and this holds here for the 34 requests for all runs except 13 and 19 (K' ≃ 33 and 35 respectively). This drop from K' ≃ 50 makes us ask, in choosing those requests with low relevant output have we simply chosen those with lower than average total output? If we raise the K' to counter-balance Run 13's low sensitivity, thus,

Request Set	Run	K	Relevant	Irrelevant	Output	K'	% Precision
34	13	87	101	1572	1672	49.18	6.04

... (VIII.2.1)

we still find Run 13 below the other runs, except Run 22, and so the poorer performance of the 34 requests in Run 13 must be due in great measure to semantic considerations. In fact, no single run at K' ≃ 50 is able to cover more than 65 out of the 93 requests at the level of 5 relevant documents, while there are 83 requests for which 5 or more are known.

We may use $\frac{K'}{K}$ as our measure of sensitivity (V.7). Runs may be rank ordered according to the cut-off values K needed to yield a given average output K' (Appendix B5), a high K giving low sensitivity. We deduce the following orders:

93 Requests: Run numbers in order of sensitivity

	(highest) ... (lowest)
K' ≃ 20	6 = 11 23 9 = 28 = 14 21 = 20 17 15 = 16 22 = 13 19 = 25 18
K' ≃ 30	6 = 11 9 28 23 14 21 22 15 18 = 13 19 = 17 = 16 20 25
K' ≃ 40	6 = 11 9 23 28 14 21 22 18 15 19 20 13 17 25 16
K' ≃ 50	6 = 9 11 28 23 14 21 = 22 18 19 13 15 20 17 16 25
Average rank order	6 = 11 9 28 = 23 14 21 22 15 18 13 20 19 17 16 25

Note that sensitivity is independent of any relevance judgments.

VIII.3 Run 13 with a Second Strategy compared with Run 13 at Higher Cut-Off

We have discussed above the numbers of new documents produced by our standard runs. The best of these was Run 21. Let us unite the document sets produced in runs 13 and 21, writing U for set union, thus:-

Request Set	Run	K	Relevant	Irrelevant	Output	K'	% Precision
93	13 U 21	71,62	1302	6006	7308	78.58	17.82
93	13	113-115	1185	5880	7065	75.97	16.77
93	13	116-117	1204	6185	7389	79.45	16.29

We have included the results for Run 13 which are nearest in total output. We do a little better by using 13 U 21 than by increasing the output for Run 13 alone. Thus the two strategies together are better than either of them singly. Although we have not distinguished documents so long as they are relevant, the requestor may appreciate the increased variety brought in by the association run.

The improvement is even more marked for the 34 set. '$\overline{13} \cap 15$' indicates documents in Run 15 but not in Run 13.

Request Set	Run	K	Relevant	Irrelevant	Output	K'	% Precision
34	13	71	79	1058	1137	33.44	6.95
34	15	73	141	1464	1605	47.20	8.79
34	13 ∩ 15	71,73	77	958	1035	30.45	7.44
34	13 U 15	71,73	156	2016	2172	63.88	7.18
34	13	110	123	1942	2065	60.74	5.96
34	13	111	124	2148	2272	66.82	5.46

Note that when two runs, at say $K' \approx 50$, are united the new average output per request will be something less than 100, the sum of their K', and vary somewhat between different pairs of runs, according to their overlap. Thus greater total relevant output may not correspond strictly with greater precision. We may compare the following figures with 13 U 21 above.

Request Set	Run	K	Relevant	Irrelevant	Output	K'	% Precision
93	13 U 19	71,66	1209	5219	6428	69.12	18.81
93	13	104-109	1158	5082	6240	67.10	18.56
93	13	110	1166	5276	6442	69.27	18.10

VIII.4 Generality of a Request

We define

$$\text{Generality of a Request} = \frac{\text{Number of Relevant Documents in Collection}}{\text{Total Number of Documents in Collection}} \quad \ldots \text{(VIII.4.1)}$$

It may be thought of as the precision should the whole library be output. Requests with low generality are in some sense likely to be harder to answer. Some previous workers who have had the whole of their document collection assessed have therefore studied generality as an additional variable. If we use total known relevant in the numerator of (VIII.4.1), the generality of our requests ranges from .0086% to .73%. We may divide our set of requests into four groups of increasing generality.

	Known Relevant	% Generality	Number of Requests
GEN 1	0-8	.000-.069	26
GEN 2	9-17	.078-.15	21
GEN 3	18-29	.16 -.25	23
GEN 4	34-84	.26 -.73	23

See Appendix B12 for the corresponding performance figures. The figures for lowest generality are perhaps too small from which to draw any strong conclusions, but the other three show a rank order very similar to those for all 93 requests together (VIII.2).

Rank Order of Run Numbers, $K' \approx 50$, for Relevant Documents

	(most)												(fewest)			
Request Set GEN 1	11	22	6	18	13	9 =	20	17		16	23	15	19	21	14
GEN 2	9	13	21	19	22	11	15	18 =		23	20	6	14	16	17
GEN 3	22	13	11	9 =	19	21	23	15		14	20	6	18	17	16
GEN 4	13	9	19	23	14	21	20	18		16	17	22	15	11	6

New Relevant Documents, not in Run 13

Request Set GEN 1	11	18	6	21	15	17	16	= 23	14	= 20	9	22	19
GEN 2	21	15	18	11	23	= 20	19	9	14	6	16	= 17	22
GEN 3	11	= 15	23	21	6	14	18	9	19	20	16	= 17	22
GEN 4	18	15	21	14	19	16	17	11	23	6	20	9	22

VIII.5 Underlining (55 requests)

Underlining (VI.2.6) was first tried with descriptor Run 14 to give it additional precision, making Run 21. Later it was tried with Runs 13 and 11 after the standard runs had been assessed. The strategies were first applied to all 93 requests at $K' \simeq 30$, and then the 55 requests selected. The figures in Appendix B13 include the resultant effective K', which do not differ enough from 30 to affect the argument.

(i) One Underlined Strategy

Referring to Appendix B13 , we see that underlining slightly improves the performance of each of the three runs so that 25 (U13) is the best of all. But U13 has quite a striking decrease in sensitivity over Run 13, needing, for example, $K = 81$ to produce $K' \simeq 50$.

(ii) Two Strategies in Succession, One Underlined

Appendix B13 shows that it is best to combine with Run 25 not Run 13 or another underlined strategy (Run 21), but an association or descriptor run, with Run 23 heading the list. Thus:-

Request Set	Run	K	Relevant	Irrelevant	Output	K'	% Precision
55	25	29	356	660	1016	18	35
55	$\overline{25} \cap 23$		155	629	784		20
55	25 U 23		511	1289	1800		28
55	25	58	>511		1853	34	>29

It would seem that Run 23 is the best to combine with Run 25, but that we could get similar performance on average out of Run 25 alone with increased cut-off. What about the requests, although underlined, for which Run 25 does badly? If a request does badly in Run 25 = U13, it may still be worth trying Run 13 but an association or descriptor run is more likely to break new ground.

However, in underlining, we are limiting ourselves to a particular subset of abstracts and it is for the requestor to decide whether this is desirable in any particular case.

IX Summary and Conclusions

IX.1 Evaluation criteria

Our aim was to find what improvements, if any, could be made upon the performance of retrieval by matching key-word stems by the use of word associations and clusters. Thus run 13 (KWS) is our standard of comparison. In assessing different strategies we consider three measures to be particularly useful:

(i) The total number of relevant documents retrieved for a set of requests in a given quantity of output.

(ii) The sensitivity of a strategy, that is, its ability to output a number o. documents uniformly close to that requested.

(iii) The coverage of a strategy, that is, the number of requests for which it retrieves some minimum number of relevant documents.

For associative strategies (i) and (iii) are calculated using total numbers of relevant documents, and also using numbers of new relevant documents with respect to the output of run 13.

As far as we are aware the concept of the sensitivity of a strategy has not previously been discussed in the literature. We measure it (V.7) as K'/K, the output per request averaged over a set of requests divided by the desired output. In changing the strategy we find that as the relative spread of the outputs for different requests about their average increases so K'/K decreases. With very low spread K'/K approaches unity. Both these effects depend on the number of strata into which the strategy partitions the output. Low sensitivity implies that the performance for some requests may suffer by very little output being produced. As regards (i) we have offset the effect of differing sensitivity by testing each strategy at the same K' rather than at the same K.

The recall performance of the various strategies may be assessed in terms of the total numbers of relevant documents they retrieve in relation to the total number of known relevant. Results of virtually exhaustive searches on three requests and of subject index searches on 12 (section VII.4) suggest that the numbers of known relevant (total and for individual requests) represent about 80% of all relevant documents in the collection.

IX.2 Performance of key-word stem searches

In relation to results yielded by our various associative strategies it must be concluded that retrieval by the simple means of comparing key-word stems (run 13) provides a very good level of performance. The total of 991 relevant documents retrieved for the full set of 93 search requests was matched only by run 28 which uses a refinement of the word stem search.* This represents a recall of about 40% of all relevant documents in the collection. Between them, all the strategies described (including run 13) produced 2,020 relevant documents.

- -

*This superiority of simple key-word stem searching when judged purely on the basis of total numbers of relevant documents retrieved closely parallels some of the findings of Cleverdon et al [15]. Our use of key-word stems is very similar to their language consisting of single terms with confounding word forms, which they found to perform better than others they tested.

The weakness of key-word stem retrieval is its very low sensitivity ($K'/K \approx 0.7$). A five word request, for example, splits the collection into at most five strata whether it contains 10,000 or 100,000 documents. There is a large spread in quantity of output with a marked tendency towards lower output than desired. Thus, in run 13 it was necessary to retrieve as close as possible to 71 documents for each request ($K = 71$) in order to produce an average output of 50 documents per request ($K' = 50$). Even at this output level run 13 produced little or no relevant material for many requests.

In run 28 the key-word stem strategy is refined by ordering the documents in each of its strata according to their coordination values in run 14. This subdivision of the strata produces excellent sensitivity ($K'/K = 0.97$). The total number of documents retrieved for 93 requests is the same as in run 13. However, this reflects a balance between 155 new relevant documents for requests for which run 13 produced low output and 155 fewer for a smaller number of requests for which run 13 produced high output. Appendix B11 summarizes results for a subset of 34 requests for each of which run 13 yielded fewer than five relevant answers. For these, run 28 produced 130 relevant documents compared with 79 in run 13.

IX.3 Advantages of associative strategies

Our strategies employ word associations in one of two ways:

(i) Word clusters formed from them are used as descriptors which are assigned to abstracts and requests. Retrieval is then based upon comparison of clusters.

(ii) Abstracts alone, or abstracts and requests, are expanded by adding associated words. Word stems in the expanded texts are then compared when retrieving.

We have found strategies of both kinds with good levels of performance and having three things to recommend them:

(a) Very good sensitivity ($K'/K > 0.9$).

(b) Fairly good recall of new relevant material. Although they retrieve fewer relevant documents in total than run 13, the best strategies add to the number of relevant documents found in that run by about 30%.

(c) Particularly good recall of new relevant material for requests for which a word stem search yields little or no relevant output. For the subset of 34 such requests the best strategies increase the number of relevant documents found in run 13 by about 75% (see Appendices B10 and B11). This improvement is accounted for partly by the better sensitivity of the associative strategies and the fact that a large proportion of the subset of 34 requests obtained low total output in run 13. However, in producing the figures in table VIII.2.1 we have compensated for this by considering a cut-off level which raises the average output of these requests to our standard of approximately 50. The figures show the improved performance of a word stem search at this output level. The best associative strategies are still able to improve upon these new figures by about 40%.

Considered individually, none of these improvements is spectacular and it would not be sensible to recommend the use of one or more associative

strategies rather than performing a simple word stem search. However, the combination of these advantages together with slightly improved coverage provided by most of the associative strategies does make their use well worth considering seriously. We think their main value lies in their use to supplement a key-word stem search when the performance of the latter is found unacceptable or when better sensitivity is required.

One most interesting observation applies generally to our associative strategies. We have found a marked negative correlation between the amount of new relevant output (relative to the results of the word stem search of run 13) from strategies and the total relevant output they produce. In other words, associative strategies providing the highest relevant output tend to produce largely the same relevant documents as run 13, those providing a good selection of new relevant material produce more meagre total numbers of relevant documents.

IX.4 Best associative strategies

Assuming that they would be used along with a key-word stem search, our final assessment of the associative strategies is based upon their performance for the subset of 34 requests for each of which Run 13 produced fewer than five relevant documents. Our 14 main strategies are ranked below according to the three criteria listed in section IX.1.

Strategies	21	15 =	18	23	11	9	14	28	6	20	17	16	19	22
New Relevant Documents+ (34 Requests)	78	77	77	70	69	64	59	54	49	46	40	39	33	17

Strategies	18	11	23 =	6	15	21	9	14	17 =	16	20	28	19	22
Coverage (34 Requests) (1 or more new relevant documents)	23	22	21	21	20	19	17	16	15	15	14	12	9	8

Strategies in average order of sensitivity	6	11	9	28	23	14	21	22	15	18	13	20	19	17	16	25
Compare: Values of K needed to give K' ≈ 50 (93 Requests)	49	50	49	52	53	55	62	62	73	65	71	73	66	76	79	81

Although no single strategy comes out on top on all counts we note that strategies 11 and 23 show a fairly balanced performance, appearing within the top six of each rank ordering. We regard these as the best of the associative strategies we have tried. Strategy 11 uses a set of 1,178 highly overlapping word groups as descriptors with proportional representation (VI.1.6). In strategy 23 key-word stems are compared after expanding both requests and abstracts by adding words which, according to connection matrix G4 (III.10), are associated with the original words.

IX.5 Effects of cluster overlap and representation

One important feature of clusters is their degree of overlap, and it is closely related to the number of words that do not become isolated in the clustering process and the number and size of the clusters themselves. In general, we find that new relevant output, coverage and sensitivity, particularly the latter, are all improved by increasing the overlap of clusters used as descriptors. By contrast, non-overlapping clusters (e.g. strategy 16) necessarily lead to lower sensitivity even than key-word stem retrieval. Use

---

+In comparing these totals with those for the 93 requests, note that 5 or more relevant documents are known to exist for only 24 out of the 34 requests.

of representation with clusters (VI.1.5) further improves these measures provided there is a sufficient degree of overlap. Proportional representation (VI.1.6) performs even better.

IX.6 Effect of obligatory request words

The results produced by allowing people to make some of their request words obligatory (indicated by underlining) in retrieved abstracts suggest that, used with restraint, this can be a useful precision device. The sensitivity of a strategy is, however, lowered.

IX.7 Effect of request generality

The generality of our requests (i.e. ratio of number of documents in collection relevant to a request to total number in collection) ranges from 0.0086% to 0.73%. These figures, however, are based upon numbers of known relevant documents, estimated to be 80% of the totals. The true generality range is therefore likely to be 0.01% to 0.91%. The 93 requests were arranged into four subsets of roughly equal size each covering a different range of generalities. The performance of different strategies was studied for these subsets of requests (VIII.4 and Appendix B12). There is no evidence of significant performance differences for the different generality ranges, nor of significantly different strategy rankings according to their performance for these classes of request.

IX.8 Choice of strategy

If it is desired to use only one strategy then a simple key-word stem search is the best when its low sensitivity is acceptable, and is certainly cheaper to implement than any associative strategy. When better sensitivity is essential, for better control of quantity of output, then a refined key-word stem search, such as run 28*, should be employed. This removes the likelihood of some requests receiving little or no output and, hence, little or no relevant output.

If better recall and sensitivity are required than it, alone, can provide a key-word stem search should be followed by an associative strategy. Of the ones we have tried strategy 23, making direct use of a set of word associations, is undoubtedly the best choice. Run 28, mentioned above, is too nearly related to Run 13 to be so useful as a second strategy. Strategy 11, though yielding a similar level of performance, would be a costlier choice requiring, as it does, the generation of a set of word clusters. We would say, more generally, that, of the strategies we have tested, those involving direct use of word associations perform comparably with, and are therefore preferable to, those employing word clusters.

If a key-word stem search is found not to be very productive, greater improvement in recall can be expected by proceeding with an associative strategy than by changing the cut-off level in the word stem search to yield the corresponding combined output.

For some of the factors affecting the cost of the various strategies, see Appendix B6.

---

*Any associative strategy can be used to subdivide the coordination strata of the word stem search. The strategy of run 23 is probably the simplest and best choice.

IX.9 Application to on-line, interactive systems

Some of the techniques we have studied could be employed to improve the performance of retrieval systems working in an on-line, interactive mode. Words and word clusters statistically associated with those in a request or in an already retrieved relevant document could be displayed. The searcher could then decide which associated words are useful in the context of his own request, and how to use them to broaden or modify his request. It seems likely that word associations could lead to greater improvements in retrieval performance when tailored in this way to a searcher's needs than when employed in a non-interactive mode.

Acknowledgements

We would like to thank Mr. R.J. Reason for sustained programming effort, for many all-night sessions on the computers and for his work on word clustering, and Mr. D.S. Baker for writing KDF9 programs needed for the retrieval tests, most of which worked first time. We are grateful to Dr. A.R. Meetham for many discussions and contributions to the work and for providing us with a set of word clusters. We would also like to thank Dr. S. Papert for many contributions to the earlier parts of the work and Mrs. V. Hawtree for her painstaking key-punching of the abstract texts.

We are indebted to the subjects in our retrieval experiments for providing us with search requests and for their conscientious assessment of the relevance of the reams of output they received. We owe special thanks to Dr. H. Rishbeth for his part as a subject and for the willing way in which he organized coverage of our geophysical material at the Radio and Space Research Station, Slough, for being a subject and for enlisting similar help from Dr. W.C. Bain, Dr. D.A. Bryant, Dr. D.L. Croom and Dr. E. Dunford. The other subjects were our colleagues in Computer Science Division at NPL - Mr. F.M. Blake, Dr. P.A.N. Briggs, Mr. D.O. Clayden, Mr. C.H. Davies, Dr. W.F. Fincham, Mr. P.H. Hammond, Mr. A.A. Hill, Mr. J. McDaniel, Mr. E.A. Newman, Dr. J.R. Parks, Mr. P.J. Pobgee, Dr. W.L. Price, Dr. P.R. Stuart, Mrs. M. Vaswani, Mr. R.S. Watson and Dr. D.M. Yates.

Figures IV.6.2 and 3 are reproduced by permission from proceedings, IFIP Congress, 1968, North-Holland Publishing Company, Amsterdam.

APPENDIX AI

WORD-STEM DICTIONARY

WITH MAXIMUM EXTENSION OF STEM LENGTH (II·5) AND FREQUENCY IN 12288 ABSTRACTS

N·B· REPETITIONS WITHIN AN ABSTRACT ARE NOT COUNTED

STEM NO	STEM	MAX EXTN	FREQ
653	ABSOR	8	384
0	ACCELERA	8	55
484	ACCESS	8	18
485	ACCURA	8	226
1	ACOUSTIC	8	51
486	ACTION	6	95
487	ACTIVE	6	95
2	ACTIVITY	8	219
654	ADAPT	8	32
962	ADD	8	171
825	ADHE	8	5
3	ADIABATI	8	9
488	ADJUST	8	86
4	ADMITTAN	8	65
489	ADVANC	8	40
5	ADVANTAG	8	93
490	AERIAL	8	150
6	AFTERNOO	8	13
963	AIR	3	147
655	ALIGN	8	36
656	ALLOY	8	32
7	ALPHANUM	8	3
8	ALTERNAT	8	131
826	ALTI	8	127
329	AMBIENT	8	20
657	AMMON	8	21
491	AMPLIF	8	1292
9	AMPLITUD	8	388
10	ANALOGUE	8	216
330	ANALOGY	7	24
658	ANALY	8	1372
11	ANALYSER	8	67
659	ANGLE	6	163
331	ANGULAR	8	75
12	ANISOTRO	8	71
332	ANNULAR	8	13
827	ANOD	6	128
492	ANOMAL	8	123
333	ANTENNA	8	23
13	APPARATU	8	89
14	APPARENT	8	77
15	APPROXIM	8	401
964	ARC	6	63
493	ARCTIC	6	17
828	AREA	5	86
16	ARGUMENT	8	19
660	ARRAY	8	35
17	ARTIFICI	8	133
494	ASPECT	6	13
18	ASSOCIAT	8	304
661	ASTRO	8	101
19	ASYMMETR	8	68
20	ASYMPTOT	8	27
21	ATMOSPHE	8	564
829	ATOM	5	57
495	ATOMIC	6	63
22	ATTACHME	8	36
23	ATTENUAT	8	172
496	AURORA	8	323
24	AUTOMATI	8	128
25	AVALANCH	8	21
334	AVERAGE	8	114
830	AXES	4	12
662	AXIAL	8	50
26	BACKGROU	8	35
27	BACKWARD	8	12
497	BALANC	8	50
28	BALANCED	8	42
[illegible]5	BALLOON	8	26
[illegible]1	BAND	5	548
[illegible]9	BANDWIDT	8	226
[illegible]6	BARRIER	8	27
832	BASE	5	641
663	BASIC	5	182
664	BASIS	5	147
965	BAY	4	33
833	BEAM	5	306
337	BEARING	8	21
834	BELT	5	71
835	BIAS	7	84
30	BIDIRECT	8	10
31	BIFURCAT	8	14
665	BINAR	8	95
32	BISTABLE	8	43
966	BIT	4	29
666	BLOCK	5	31
33	BLOCKING	8	33
967	BOD	6	68
338	BOMBARD	8	25
836	BOND	8	13
339	BOUNDAR	8	104
498	BRANCH	8	37
34	BREMSSTR	8	6
499	BRIDGE	7	108
500	BRIGHT	8	82
667	BURST	8	88
35	CALCULAT	8	899
36	CALIBRAT	8	27
340	CANONIC	8	10
341	CAPACIT	8	468
501	CARBON	8	40
342	CARRIER	8	116
502	CASCAD	8	71
503	CATHOD	8	250
668	CAVIT	8	239
837	CELL	8	88
37	CENTIMET	8	55
343	CENTRAL	7	42
504	CENTRE	7	147
38	CERENKOV	8	13
344	CHANNEL	8	70
39	CHARACTE	8	921
669	CHARG	8	383
838	CHOP	8	20
670	CIRCL	8	31
345	CIRCUIT	8	1863
40	CIRCULAR	8	129
41	CIRCULAT	8	37
671	CLASS	8	123
346	CLASSIC	8	61
347	CLASSIF	8	42
505	CLIMAT	8	8
506	CLOSED	8	43
672	CLOUD	8	91
839	COAT	7	34
840	COAX	8	113
968	COD	6	33
42	COEFFICI	8	318
507	COHERE	6	55
841	COIL	7	112
43	COINCIDE	8	57
842	COLD	4	47
348	COLLECT	8	114
44	COLLISIO	8	151
508	COLOUR	8	17
509	COLUMN	8	59
510	COMBIN	8	188
45	COMMUNIC	8	65
46	COMPARIS	8	202
47	COMPATIB	8	7
48	COMPENSA	8	119
49	COMPLEME	8	19
50	COMPONEN	8	416
349	COMPLEX	8	151

STEM NO	STEM	MAX EXTN	FREQ
511	COMPOS	8	85
51	COMPRESS	8	27
512	COMPUT	8	559
52	CONCENTR	8	115
53	CONDENSE	8	14
54	CONDITIO	8	565
350	CONDUCT	8	569
843	CONE	5	20
55	CONFIGUR	8	67
56	CONFORMA	8	12
673	CONIC	8	9
57	CONJUNCT	8	45
674	CONNE	8	246
351	CONSERV	8	15
352	CONSTAN	8	472
58	CONSTITU	8	51
353	CONTACT	8	147
354	CONTINU	8	262
355	CONTOUR	8	28
356	CONTROL	8	472
59	CONVENTI	8	87
60	CONVERSI	8	71
357	CONVERT	8	152
513	CONVEX	8	10
844	COOL	8	31
845	CORE	5	204
514	CORONA	8	86
358	CORPUSC	8	54
359	CORRECT	8	164
61	CORRELAT	8	267
62	CORRESPO	8	228
63	CORRUGAT	8	5
515	COSMIC	6	267
675	COUNT	8	178
676	COUPL	8	401
64	CRITICAL	8	199
677	CROSS	8	178
65	CRYOTRON	8	9
360	CRYSTAL	8	328
361	CURRENT	8	842
846	CURV	8	381
678	CYCLE	8	115
66	CYCLOTRO	8	37
516	CYLIND	8	221
679	DAILY	5	55
680	DAMPI	8	70
847	DARK	8	19
848	DATA	4	508
969	DAY	4	192
362	DAYTIME	7	58
997	DB	2	31
998	DC	5	221
681	DECAY	8	80
363	DECIMAL	8	29
364	DECIMET	8	16
682	DEFIN	8	169
517	DEFORM	8	19
518	DEGREE	8	102
683	DELAY	8	224
67	DEMODULA	8	30
684	DENSE	8	21
519	DENSIT	8	597
68	DEPARTUR	8	25
365	DEPOSIT	8	23
685	DEPTH	5	32
686	DERIV	8	816
69	DERIVATI	8	83
520	DESIGN	8	1035
521	DETECT	8	253
366	DETONAT	8	8
367	DEVELOP	8	504
522	DEVIAT	8	78
523	DIAMET	8	93
70	DIELECTR	8	247
71	DIFFEREN	8	726
72	DIFFRACT	8	215
524	DIFFUS	8	165
687	DIGIT	8	253
688	DILUT	8	11
73	DIMENSIO	8	167
689	DIODE	8	345
690	DIPOL	8	115
525	DIRECT	8	269
74	DIRECTIO	8	222
75	DIRECTIV	8	9
970	DIS	4	125
76	DISCHARG	8	345
77	DISCONTI	8	45
78	DISCRETE	8	49
368	DISPERS	8	103
369	DISPLAC	8	44
79	DISSIPAT	8	64
526	DISTAN	8	176
370	DISTORT	8	185
80	DISTRIBU	8	688
371	DISTURB	8	289
372	DIURNAL	7	204
527	DIVERG	8	14
849	DIVI	8	134
528	DOMAIN	8	29
373	DOPPLER	7	75
691	DOUBL	8	151
850	DRAG	4	15
851	DRAW	8	33
692	DRIFT	8	237
852	DRIV	7	120
853	DROP	8	47
854	DRUM	5	39
855	DUAL	8	29
856	DUCT	8	9
374	DYNAMIC	8	95
529	DYNAMO	7	44
693	EARTH	8	506
857	EAST	4	37
81	ECCENTRI	8	22
858	ECHO	8	290
530	ECLIPS	8	97
971	EDD	7	17
859	EDGE	4	45
82	EFFECTIV	8	191
83	EFFICIEN	8	113
84	ELECTRIC	8	703
375	ELECTRO	8	123
85	ELECTROD	8	176
86	ELECTROM	8	466
87	ELECTRON	8	1642
88	ELECTROS	8	51
376	ELEMENT	8	454
694	ELLIP	8	73
531	EMISSI	8	380
860	EMIT	8	71
377	EMITTER	8	94
532	ENERGY	6	489
89	ENGINEER	8	40
695	EQUAL	8	123
378	EQUALIZ	8	40
696	EQUAT	8	603
379	EQUATOR	8	142
90	EQUILIBR	8	65
91	EQUIVALE	8	258
697	ERROR	6	179
698	ERUPT	8	22
380	EVALUAT	8	136
92	EVAPORAT	8	18
699	EXACT	5	88
533	EXAMIN	8	315
93	EXCHANGE	8	43
94	EXCITATI	8	127
381	EXCITED	7	76
382	EXCITON	8	12
95	EXOSPHER	8	34
700	EXPAN	8	77
383	EXPLOSI	8	40
96	EXPONENT	8	59
534	EXTEND	8	204
97	EXTENSIO	8	120

STEM NO	STEM	MAX EXTN	FREQ
98	EXTERNAL	8	130
99	EXTINCTI	8	5
100	EXTRAORD	8	29
535	EXTREM	8	50
861	FACE	5	22
536	FACTOR	8	282
537	FADING	6	75
862	FAIL	8	29
863	FALL	8	57
972	FAR	3	49
384	FARADAY	8	52
864	FAST	4	87
701	FAULT	6	10
865	FEED	7	43
101	FEEDBACK	8	340
702	FERRI	8	22
385	FERRITE	8	144
102	FERROELE	8	26
103	FERROMAG	8	128
703	FIELD	6	1646
704	FIGUR	8	128
104	FILAMENT	8	28
866	FILM	5	145
867	FILT	8	475
868	FINE	4	28
538	FINITE	6	93
973	FIX	8	85
705	FLARE	6	149
706	FLASH	8	2[illegible]
539	FLIGHT	6	[illegible]
869	FLIP	4	[illegible]
870	FLOW	4	84
105	FLUCTUAT	8	211
707	FLUID	8	37
871	FLUX	8	142
974	FOC	8	59
708	FORCE	5	94
872	FORM	4	335
106	FORMATIO	8	75
386	FOURIER	7	60
107	FRACTION	8	39
873	FREE	4	209
108	FREQUENC	8	2146
109	FUNCTION	8	712
110	FUNDAMEN	8	116
874	GAIN	4	321
111	GALACTIC	8	53
709	GALAX	8	38
975	GAP	3	41
976	GAS	3	359
387	GASEOUS	7	35
710	GASES	5	105
977	GAT	8	59
711	GAUSS	8	14
112	GAUSSIAN	8	20
388	GENERAT	8	458
113	GEOGRAPH	8	29
114	GEOMAGNE	8	466
389	GEOMETR	8	83
115	GEOPHYSI	8	58
116	GERMANIU	8	40
712	GIANT	6	7
713	GLASS	5	21
117	GRADIENT	8	73
714	GRAPH	8	289
540	GRAVIT	8	51
875	GRID	5	144
541	GROUND	8	172
715	GROUP	8	108
542	GROWTH	6	29
876	GYRO	8	50
118	HAMILTON	8	18
119	HARMONIC	8	220
877	HEAD	5	22
878	HEAT	4	55
543	HEATER	7	21
390	HEATING	7	55
[illegible]16	HEAVY	5	20
544	HEIGHT	7	472
717	HELIC	8	17
545	HELIUM	6	28
120	HEMISPHE	8	54
879	HIGH	7	1265
880	HISS	7	5
881	HOLE	5	35
121	HOMOGENE	8	61
122	HORIZONT	8	115
882	HOUR	6	126
546	HYBRID	8	25
123	HYDRODYN	8	25
124	HYDROGEN	8	81
125	HYDROMAG	8	56
126	HYPERBOL	8	18
127	HYPOTHES	8	63
718	IDEAL	8	103
547	IDENTI	8	109
978	IDL	6	12
719	IMAGE	6	69
548	IMPACT	7	26
128	IMPEDANC	8	392
129	IMPERFEC	8	27
549	IMPULS	8	35
720	IMPUR	8	25
130	IMPURITY	8	28
391	INCIDEN	8	244
550	INCLIN	8	21
131	INCOHERE	8	17
132	INCREMEN	8	12
133	INDEPEND	8	168
721	INDEX	5	113
392	INDICAT	8	601
722	INDUC	8	61
134	INDUCTAN	8	107
393	INDUCTI	8	74
135	INDUCTOR	8	42
394	INERTIA	8	18
395	INFINIT	8	123
136	INFORMAT	8	119
137	INHOMOGE	8	66
396	INITIAL	8	95
551	INJECT	8	25
723	INNER	5	40
724	INPUT	6	460
138	INSERTIO	8	62
139	INSTABIL	8	51
140	INSTRUME	8	108
397	INSULAT	8	39
141	INTEGRAL	8	122
142	INTEGRAT	8	119
143	INTENSIT	8	288
144	INTERACT	8	219
145	INTERFER	8	144
146	INTERMED	8	37
147	INTERNAL	8	83
148	INTERNAT	8	62
149	INTERPLA	8	52
150	INTERVAL	8	57
151	INVARIAN	8	20
398	INVERSE	8	61
152	INVERSIO	8	34
552	INVERT	8	42
979	ION	4	312
725	IONIC	5	57
726	IONIZ	8	248
153	IONIZATI	8	355
154	IONOGRAM	8	34
155	IONOSOND	8	18
156	IONOSPHE	8	1 88
157	IRREGULA	8	142
158	IRREVERS	8	9
553	ISLAND	6	25
554	ISOLAT	8	42
159	ISOTHERM	8	12
399	ISOTROP	8	38
555	ITERAT	8	29
727	JOINT	8	8

STEM NO	STEM	MAX EXTN	FREQ
883	JUMP	8	13
160	JUNCTION	8	199
400	JUPITER	7	26
401	KINETIC	7	51
161	LABORATO	8	55
556	LADDER	7	76
980	LAG	4	30
884	LAND	4	14
162	LATITUDE	8	253
402	LATTICE	8	102
557	LAUNCH	8	33
981	LAW	4	111
728	LAYER	6	613
885	LEAD	4	61
403	LEADING	7	37
729	LEAOS	5	63
886	LEAK	8	29
558	LENGTH	8	133
730	LEVEL	8	315
887	LIFE	8	63
731	LIGHT	6	136
163	LIGHTNIN	8	64
732	LIMIT	8	301
888	LINE	5	490
559	LINEAR	8	454
560	LIQUID	8	52
889	LOAD	8	217
733	LOCAL	5	109
734	LOCAT	8	101
890	LOCK	8	37
735	LOGIC	8	84
891	LONG	4	194
164	LONGITUD	8	97
892	LOOP	5	157
893	LOSS	6	292
982	LOW	5	971
165	LUMINESC	8	27
983	LUN	5	86
404	MACHINE	8	74
561	MAGNET	7	24
405	MAGNETI	8	88
166	MAGNETIC	8	1571
167	MAGNETIS	8	24
406	MAGNETO	7	62
168	MAGNETOM	8	26
169	MAGNETOR	8	13
170	MAGNETOS	8	48
171	MAGNITUD	8	118
894	MAIN	4	129
736	MAJOR	5	44
737	MASER	6	237
895	MASK	8	10
896	MASS	4	80
738	MATCH	8	83
172	MATHEMAT	8	155
739	MATRI	8	154
562	MATTER	6	22
740	MAXIM	8	419
897	MEAN	4	194
563	MEASUR	8	1294
173	MECHANIC	8	160
174	MECHANIS	8	212
741	MEDIA	5	67
564	MEDIAN	7	17
565	MEDIUM	7	160
742	MEMOR	8	80
566	MERCUR	8	31
743	METAL	8	321
567	METEOR	8	199
744	METRE	6	31
175	MICROMIN	8	11
176	MICROPUL	8	24
177	MICROWAV	8	395
178	MIDNIGHT	8	32
179	MILLIMET	8	44
180	MILLIMIC	8	14
181	MINIATUR	8	33
745	MINIM	8	256

STEM NO	STEM	MAX EXTN	FREQ
568	MIRROR	8	30
984	MIX	8	70
746	MOBIL	8	26
898	MODE	5	213
747	MODEL	6	287
748	MODUL	7	19
407	MODULAT	8	306
408	MOLECUL	8	147
569	MOMENT	7	50
182	MOMENTUM	8	29
183	MONOSTAB	8	20
749	MONTH	7	8[illegible]
899	MOON	5	65
184	MORPHOLO	8	14
570	MOTION	8	224
750	MOTOR	6	25
185	MOVEMENT	8	108
571	MOVING	6	109
186	MULTILAY	8	13
409	MULTIPL	8	102
187	MULTIPLI	8	125
188	MULTIPOL	8	14
189	MULTISTA	8	40
190	MULTIVIB	8	72
572	NARROW	6	80
410	NATURAL	7	58
573	NEBULA	7	32
411	NEGATIV	8	319
985	NET	4	26
412	NETWORK	8	649
413	NEUTRAL	8	92
414	NEUTRON	8	36
751	NIGHT	5	119
191	NITROGEN	8	38
900	NOIS	5	546
192	NONLINEA	8	336
193	NONMAGNE	8	8
194	NONPERIO	8	2
195	NONRECIP	8	33
196	NONSEASO	8	3
197	NONUNIFO	8	45
574	NORMAL	6	165
198	NORMALIZ	8	25
752	NORTH	8	95
901	NOTE	5	337
415	NUCLEAR	8	102
575	NUMBER	7	398
199	NUMERICA	8	143
576	OBLIQU	8	56
200	OBSERVAT	8	878
986	OIL	8	9
902	OPEN	4	30
201	OPERATIO	8	490
753	OPTIC	7	178
754	OPTIM	8	135
755	ORBIT	6	124
416	ORBITAL	8	45
756	ORDER	8	297
577	ORIENT	8	36
578	ORIGIN	6	108
202	ORTHOGON	8	20
203	OSCILLAT	8	928
204	OSCILLOG	8	29
205	OUTBURST	8	18
757	OUTER	5	89
579	OUTPUT	7	583
417	OVERSHO	8	15
903	OXID	6	53
580	OXYGEN	8	53
904	PAIR	8	100
758	PAPER	6	246
418	PARABOL	8	71
206	PARALLEL	8	304
207	PARAMAGN	8	82
419	PARAMET	8	521
420	PARTIAL	8	78
208	PARTICLE	8	[illegible]4
905	PASS	4	[illegible]0

STEM NO	STEM	MAX EXTN	FREQ
421	PASSAGE	8	55
581	PASSIV	8	79
906	PATH	5	142
422	PATTERN	8	104
907	PEAK	7	155
209	PENETRAT	8	39
423	PENTODE	8	47
210	PERFORMA	8	260
424	PERIGEE	7	14
582	PERIOD	8	441
211	PERMANEN	8	23
425	PERMEAB	8	42
212	PERPENDI	8	64
426	PERSIST	8	21
427	PERTURB	8	116
908	PHAS	7	529
213	PHENOMEN	8	313
909	PHOT	5	38
214	PHOTOCEL	8	14
215	PHOTOCON	8	13
216	PHOTOELE	8	44
217	PHOTOGRA	8	64
583	PHOTON	7	22
218	PHOTOSEN	8	4
584	PHYSIC	8	179
428	PICTURE	8	14
219	PIEZOELE	8	34
585	PLANAR	6	12
759	PLANE	6	324
586	PLANET	8	36
587	PLASMA	7	458
429	PLASTIC	8	19
760	PLATE	6	115
910	PLOT	8	77
761	POINT	6	325
762	POLAR	5	109
220	POLARITY	8	15
430	POLARIZ	8	192
911	POLE	5	109
221	POLYNOMI	8	35
222	POPULATI	8	11
912	PORT	5	25
223	POSITION	8	145
224	POSITIVE	8	214
225	POTENTIA	8	218
226	POTENTIO	8	25
763	POWER	6	661
588	PRECIS	8	65
431	PREDICT	8	136
589	PRESET	8	4
227	PRESSURE	8	200
432	PRIMARY	7	99
228	PRINCIPA	8	45
764	PRINT	8	64
229	PROBABIL	8	39
765	PROBE	6	80
913	PROD	8	563
433	PRODUCT	7	33
434	PROFILE	8	63
435	PROGRAM	8	93
230	PROGRESS	8	35
231	PROPAGAT	8	375
232	PROPORTI	8	125
436	PROTECT	8	19
590	PROTON	8	90
914	PULL	7	96
591	PULSAT	8	34
766	PULSE	6	640
915	PUMP	8	75
916	PURE	4	32
233	QUADRIPO	8	138
234	QUADRUPD	8	14
437	QUALITY	7	38
767	QUANT	8	207
768	QUIET	5	54
769	RADAR	6	166
592	RADIAL	8	40
438	RADIANT	7	11

STEM NO	STEM	MAX EXTN	FREQ
593	RADIAT	8	57
235	RADIATIO	8	609
770	RADII	5	33
771	RADIO	5	973
236	RADIOSON	8	15
594	RADIUS	6	58
595	RANDOM	8	104
917	RANG	8	667
918	RATE	4	206
772	RATIO	6	229
987	RAY	4	259
237	REACTANC	8	106
238	REACTION	8	37
439	REACTIV	8	41
440	REACTOR	8	45
919	READ	6	45
441	READING	7	24
920	REAL	4	72
596	RECEIV	8	238
239	RECIPROC	8	38
442	RECOGNI	8	17
240	RECOMBIN	8	98
597	RECORD	8	442
443	RECOVER	8	30
598	RECTAN	8	103
599	RECTIF	8	190
988	RED	3	14
773	REDUC	8	275
241	REFERENC	8	312
444	REFLECT	8	359
242	REFLECTO	8	37
243	REFLEXIO	8	7
445	REFRACT	8	128
244	REGENERA	8	38
446	REGULAT	8	69
245	RELATIVE	8	190
246	RELATIVI	8	48
247	RELAXATI	8	126
774	RELAY	6	49
447	RELEASE	8	14
248	REPRESEN	8	275
249	RESIDUAL	8	29
250	RESISTAN	8	397
251	RESISTIV	8	62
252	RESISTOR	8	155
253	RESOLUTI	8	57
600	RESOLV	8	39
601	RESONA	8	133
254	RESONANC	8	365
255	RESONATO	8	160
256	RESPONSE	8	332
257	RESTRICT	8	42
602	RETARD	8	35
603	RETURN	6	23
448	REVERSA	8	25
449	REVERSE	8	47
258	REVERSIB	8	14
775	REVOL	8	26
776	RIGID	8	15
921	RING	4	58
604	RIPPLE	7	21
922	RISE	4	105
923	ROCK	5	4
605	ROCKET	8	179
989	ROD	4	50
924	RDDM	4	28
777	ROTAT	8	188
259	ROUGHNES	8	11
778	ROUND	8	29
990	ROW	4	9
779	SAMPL	8	91
260	SATELLIT	8	435
450	SATURAB	8	32
451	SATURAT	8	75
606	SCALAR	7	35
780	SCALE	6	127
925	SCAN	8	32
452	SCATTER	8	302

STEM NO	STEM	MAX EXTN	FREQ
261	SCINTILL	8	77
607	SCREEN	8	116
608	SEARCH	8	15
609	SEASON	8	138
262	SECONDAR	8	153
453	SECTION	8	216
454	SEGMENT	8	9
610	SELECT	8	82
263	SELECTIV	8	61
781	SELEN	8	21
926	SELF	4	142
264	SEMICOND	8	152
265	SEMIDIUR	8	24
455	SENSING	7	13
266	SENSITIV	8	117
267	SEQUENCE	8	23
268	SEQUENTI	8	10
611	SERIES	6	309
782	SERVO	8	120
991	SET	4	115
783	SHAPE	6	159
456	SHAPING	7	20
784	SHARP	8	41
785	SHEAR	8	12
786	SHEET	6	51
787	SHELL	6	22
788	SHIFT	8	190
789	SHOCK	6	31
790	SHORT	8	261
927	SHOT	5	17
612	SHOWER	7	32
791	SHUNT	8	57
928	SIGN	5	24
613	SIGNAL	8	605
792	SILIC	7	63
614	SILVER	8	18
457	SIMILAR	8	226
458	SIMULAT	8	48
269	SIMULTAN	8	167
793	SINGL	6	328
270	SINUSOID	8	90
929	SIZE	4	107
930	SKIN	4	40
931	SLIT	5	28
932	SLOW	4	66
933	SOFT	4	19
794	SOLAR	5	680
795	SOLID	6	168
271	SOLUTION	8	340
796	SOLVE	6	77
272	SOUNDING	8	62
615	SOURCE	7	444
797	SOUTH	8	98
798	SPACE	5	255
799	SPARK	8	28
459	SPATIAL	8	44
273	SPECIMEN	8	21
616	SPECTR	8	316
274	SPECTROG	8	27
275	SPECTROM	8	71
276	SPECTROS	8	44
800	SPEED	6	206
617	SPHERE	7	76
277	SPHERICA	8	76
278	SPHEROID	8	13
934	SPIN	5	128
801	SPLIT	5	15
618	SPLITT	8	37
279	SPONTANE	8	23
280	SPORADIC	8	85
619	SPREAD	6	76
281	SPURIOUS	8	10
460	SPUTNIK	8	43
620	SQUARE	8	161
935	STAB	8	162
461	STABILI	8	466
462	STAGGER	8	27
282	STANDARD	8	140

STEM NO	STEM	MAX EXTN	FREQ
283	STANDING	8	22
936	STAR	5	101
802	START	8	42
621	STATIC	8	85
463	STATION	8	206
284	STATIONA	8	35
285	STATISTI	8	174
622	STEADY	6	129
937	STEP	5	105
623	STIMUL	8	23
286	STOCHAST	8	6
938	STOR	8	239
803	STORM	6	243
464	STRAIGH	8	33
465	STRATIF	8	20
624	STREAM	8	80
287	STRENGTH	8	169
625	STRESS	8	28
804	STRIP	6	48
626	STROKE	7	21
288	STRUCTUR	8	206
289	SUBMILLI	8	15
290	SUCCESSI	8	56
992	SUM	8	160
627	SUMMER	6	64
993	SUN	4	175
466	SUNRISE	7	39
628	SUNSET	8	22
467	SUNSPOT	8	178
291	SUPERCON	8	54
292	SUPERREG	8	7
468	SUPPLIE	8	62
469	SURFACE	8	338
805	SURGE	6	9
629	SURVEY	8	176
806	SWEEP	8	49
630	SWITCH	8	272
631	SYMBOL	8	17
470	SYMMETR	8	160
293	SYNCHRON	8	66
294	SYNTHESI	8	169
632	SYSTEM	7	958
633	TABULA	8	127
939	TAIL	8	9
940	TAPE	5	61
634	TARGET	7	22
295	TELESCOP	8	60
471	TELEVIS	8	55
296	TEMPERAT	8	597
297	TEMPORAL	8	10
635	TENSOR	7	54
298	TERMINAL	8	140
299	TERMINAT	8	59
300	TERRESTR	8	53
807	THEOR	8	1878
472	THERMAL	7	158
301	THERMIST	8	40
302	THERMOCO	8	8
303	THERMODY	8	24
304	THERMOEL	8	29
305	THICKNES	8	77
941	THIN	4	144
306	THRESHOL	8	38
473	THUNDER	8	51
307	THYRATRO	8	36
994	TID	5	54
942	TIME	4	751
636	TIMING	6	14
308	TOLERANC	8	25
943	TONE	5	16
637	TOROID	8	43
474	TORSION	8	14
808	TRACE	6	27
809	TRACK	8	47
810	TRAIL	8	81
811	TRAIN	6	32
309	TRAJECTO	8	39
310	TRANSDUC	8	65

STEM NO	STEM	MAX EXTN	FREQ	STEM NO	STEM	MAX EXTN	FREQ
311	TRANSFER	8	195	647	VECTOR	8	101
312	TRANSFOR	8	386	815	VEHIC	8	20
313	TRANSIEN	8	179	480	VELOCIT	8	309
314	TRANSIST	8	679	325	VERTICAL	8	218
475	TRANSIT	7	39	648	VIBRAT	8	69
315	TRANSITI	8	105	950	VIEW	8	61
476	TRANSMI	8	394	481	VIRTUAL	7	38
316	TRANSPOR	8	64	482	VISIBLE	7	40
317	TRANSVER	8	76	649	VISUAL	8	48
944	TRAP	8	78	951	VOLT	8	755
318	TRAVELLI	8	113	650	VOLUME	8	43
638	TRIANG	8	32	816	WAFER	6	9
477	TRIGGER	8	92	952	WALL	5	42
639	TRIODE	7	117	953	WASH	8	29
640	TRIPLE	6	14	817	WATER	5	41
641	TROUGH	7	11	954	WAVE	5	1154
945	TUBE	5	206	326	WAVEFORM	8	137
995	TUN	6	198	327	WAVEGUID	8	136
478	TUNABLE	7	35	328	WAVELENG	8	231
642	TUNNEL	6	63	818	WEDGE	6	28
319	TURBULEN	8	44	819	WEIGH	8	31
812	TWIST	8	7	955	WELL	5	125
320	UNIDIREC	8	18	651	WHISTL	8	78
946	UNIF	8	49	820	WHITE	5	18
479	UNIFORM	7	125	821	WIDTH	6	138
643	UNIQUE	8	28	956	WIND	5	109
947	UNIT	7	336	483	WINDING	8	56
321	UNIVERSA	8	40	652	WINTER	6	79
322	UNSTABLE	8	14	957	WIRE	5	101
813	UPPER	5	216	958	WORK	4	256
948	VACU	6	75	822	WORLD	5	70
644	VALENC	8	6	823	WOUND	5	2
645	VALLEY	7	4	959	WRIT	7	30
814	VALVE	6	380	999	X	1	85
646	VAPOUR	7	42	960	YEAR	6	155
323	VARIABLE	8	305	824	YIELD	8	58
324	VARIATIO	8	939	961	ZERO	6	169
949	VARY	7	146	996	ZON	5	131

APPENDIX A2

4-Zone Card Code

The table below shows the code used when punching the texts of abstracts on cards for input to the computer. Each character, numeric or alphabetic, is punched in a separate column of a card. Each numeral in the range 0-9 is coded by a single hole punched in the appropriate row. Each alphabetic character is coded by one hole in one of the rows Y,X,0 and another in one of the remaining rows.

ROW \ ZONE	NUMERALS	Y	X	0
0	0	-	-	-
1	1	A	B	C
2	2	D	E	F
3	3	G	H	I
4	4	J	K	L
5	5	M	N	O
6	6	P	Q	R
7	7	S	T	U
8	8	V	W	X
9	9	Y	Z	F.S.

Fig. A.2.1

APPENDIX A3

Letter Frequencies within Abstracts

The program used to read the abstracts from punched cards, perform the dictionary look-up, etc., also counted the number of occurrences of each letter of the alphabet and the number of spaces. All words appearing in the texts, not only dictionary words, were included in this frequency count. The cumulative frequencies obtained from 11,571 abstracts appear below in descending order. The average word length, as computed from these frequencies, is 5.5 letters.

Symbol	Frequency
Space	477 860
E	324 558
T	233 300
I	224 127
A	214 748
O	197 661
N	189 274
R	182 768
S	171 761
C	113 986
L	101 683
D	99 513
H	88 789
F	77 790

Symbol	Frequency
U	70 815
M	66 876
P	62 456
G	47 629
Y	33 544
B	32 970
V	29 854
W	26 040
X	7 469
Q	7 239
K	5 744
Z	3 467
J	1 295

APPENDIX A4

WORD-FREQUENCY DISTRIBUTIONS
1 BATCH = 1536 ABSTRACTS
F = FREQUENCY = NUMBER OF ABSTRACTS
N(F) = NUMBER OF WORDS WITH FREQUENCY F

BATCH 1

F	N(F)	F	N(F)
0	0	50	5
1	14	51	2
2	109	52	5
3	113	53	2
4	86	55	3
5	75	56	3
6	50	57	2
7	49	59	1
8	38	60	1
9	41	62	1
10	24	63	2
11	19	64	1
12	25	66	1
13	29	67	1
14	20	68	1
15	21	69	1
16	13	70	1
17	23	72	1
18	16	74	1
19	11	75	1
20	16	77	1
21	8	79	3
22	9	80	2
23	11	81	2
24	12	86	1
25	4	87	1
26	4	88	3
27	5	92	1
28	6	93	1
29	8	96	1
30	3	97	1
31	7	103	1
32	5	106	1
34	6	107	2
35	6	111	1
36	5	116	1
37	10	124	1
38	1	133	1
39	3	140	2
40	3	143	1
41	3	145	1
42	4	170	1
43	1	175	1
44	3	201	1
45	4	203	1
47	2	214	1
48	2	220	1
49	1	223	1

BATCHES 1-2

F	N(F)	F	N(F)
0	0	78	4
1	4	79	1
2	23	80	3
3	37	81	5
4	49	82	2
5	46	83	2
6	47	84	2
7	40	86	1
8	49	87	3
9	41	88	1
10	37	89	2
11	29	92	1
12	36	93	2
13	14	94	2
14	19	95	1
15	24	96	1
16	20	97	2
17	21	98	1
18	21	99	1
19	20	101	2
20	18	102	2
21	10	105	2
22	16	107	1
23	6	108	2
24	13	110	2
25	11	112	1
26	16	113	1
27	11	114	1
28	12	116	1
29	9	117	1
30	9	119	1
31	8	120	1
32	9	125	1
33	5	126	1
34	8	127	2
35	9	128	3
36	8	129	1
37	5	132	1
38	9	133	1
39	11	138	2
40	6	144	1
41	9	146	4
42	4	153	2
43	4	154	1
44	1	155	1
45	4	161	1
46	4	162	1
47	6	167	1
48	3	168	1
49	9	171	1
50	1	175	1
51	6	187	1
52	2	193	1
53	3	197	1
54	1	198	1
56	1	204	1
57	4	209	1
58	4	223	2
59	2	230	1
60	1	238	2
61	1	258	1
62	2	275	1
63	4	282	1
64	6	285	1
65	1	293	1
66	7	304	1
68	2	310	1
69	2	386	1
70	1	402	1
71	2	405	1
72	4	407	1
73	2	409	1
75	1	416	1
77	4		

BATCHES 1-4

F	N(F)	F	N(F)	F	N(F)
0	0	76	2	174	1
1	1	77	3	177	1
2	4	78	3	179	1
3	11	79	5	181	1
4	8	81	4	183	1
5	18	82	1	184	1
6	17	83	1	186	1
7	17	84	4	193	1
8	17	85	5	196	2
9	23	87	2	198	2
10	28	88	6	199	2
11	21	89	3	200	1
12	29	90	2	204	1
13	17	91	2	205	1
14	20	92	3	206	1
15	31	93	5	208	1
16	16	96	2	210	2
17	25	97	2	213	1
18	13	98	3	214	1
19	16	99	5	218	2
20	17	100	1	221	2
21	12	101	4	225	1
22	16	102	1	227	1
23	11	103	2	228	2
24	9	104	4	232	1
25	17	105	3	237	1
26	10	106	5	241	1
27	13	107	2	245	1
28	12	108	3	246	1
29	14	109	2	247	1
30	10	111	3	251	1
31	7	112	1	255	1
32	12	114	2	269	1
33	8	115	1	270	1
34	11	116	2	274	1
35	3	117	1	278	1
36	9	118	2	281	1
37	9	120	1	287	1
38	4	121	1	288	2
39	8	122	1	289	2
40	11	124	1	293	1
41	10	125	2	298	1
42	5	126	1	302	1
43	6	129	3	305	3
44	6	130	2	312	1
45	7	132	2	314	2
46	4	133	1	319	1
47	7	134	1	325	1
48	10	135	1	333	1
49	7	136	2	339	1
50	4	137	1	343	1
51	8	138	1	345	2
52	11	139	2	380	1
53	5	140	2	387	1
54	5	141	1	392	1
55	6	142	1	395	1
56	4	143	2	411	1
57	7	144	3	446	2
58	5	145	1	448	1
59	7	147	1	459	1
60	6	148	2	471	1
61	4	149	1	513	1
62	6	150	2	531	1
63	3	156	2	534	1
64	2	157	1	539	1
65	3	158	3	596	1
66	7	159	3	616	1
67	6	160	3	631	1
68	4	161	3	639	1
69	4	163	2	760	1
70	6	164	2	784	1
71	2	165	2	840	1
72	4	168	1	873	1
73	1	170	2	906	1
74	3	172	2	996	1
75	3	173	1		

BATCHES 1-3

F	N(F)
0	0
2	1
3	2
4	4
5	4
6	3
7	5
8	5
9	12
10	6
11	6
12	9
13	11
14	19
15	9
16	2
17	9
18	10
19	8
20	11
21	12
22	12
23	7
24	8
25	13
26	8
27	9
28	11
29	16
30	6
31	8
32	12
33	10
34	7
35	11
36	6
37	11
38	9
39	9
40	14
41	4
42	11
43	7
44	6
45	8
47	5
48	5
49	5
50	6
51	8
52	3
53	4
54	6
55	9
56	4
57	8
58	6
59	5
60	2
61	8
62	6
63	8
64	7
65	6
66	3
67	3
68	2
69	5
70	4

F	N(F)
71	7
72	2
73	2
74	2
75	7
76	6
77	6
78	4
79	2
80	6
81	2
82	3
83	4
84	3
85	5
86	5
87	2
88	4
89	2
90	2
91	2
92	2
93	4
94	2
95	6
96	1
97	2
98	2
99	1
100	1
101	5
102	4
103	3
104	3
105	4
106	1
107	2
108	5
109	6
111	1
112	1
113	4
114	2
115	6
116	4
117	2
118	1
119	4
120	5
122	1
123	5
124	1
125	5
126	2
127	4
128	6
129	3
130	1
131	2
133	3
134	1
135	1
136	4
137	1
138	3
140	2
142	5
143	1

F	N(F)
144	4
145	2
146	1
147	5
149	1
150	1
151	3
152	2
153	1
154	1
155	4
157	1
158	1
159	1
160	4
161	1
162	1
163	1
164	1
165	2
166	1
167	2
168	2
169	4
171	1
172	2
174	1
175	1
176	3
178	4
179	4
182	1
185	1
188	2
190	3
191	1
192	2
194	2
195	1
198	1
199	3
200	1
202	1
204	3
206	5
207	1
209	1
211	1
212	1
213	1
214	1
215	1
216	3
217	1
218	2
219	2
220	1
221	2
222	1
224	2
226	3
228	1
229	1
230	1
231	1
237	2
238	1
239	2

F	N(F)
243	1
244	1
246	2
247	1
248	1
250	1
253	3
255	1
256	2
258	1
259	1
260	1
261	1
262	1
267	2
269	1
272	1
275	2
282	1
287	1
288	1
289	2
290	1
292	1
297	1
301	1
302	1
304	2
305	1
306	2
309	2
312	2
313	1
314	1
315	2
316	1
318	1
319	1
321	2
323	1
324	1
325	1
328	2
332	1
335	1
336	2
337	1
338	1
340	2
345	2
355	1
359	2
365	1
375	1
380	2
381	1
383	1
384	1
386	1
388	1
392	1
394	1
395	1
397	1
398	1
401	2
416	1
419	1
435	1

F	N(F)
441	1
442	1
444	1
454	2
456	2
460	1
466	3
468	1
472	3
475	1
489	1
490	2
504	1
506	1
508	1
521	1
529	1
546	1
548	1
559	1
563	1
564	1
565	1
569	1
585	1
597	2
601	1
603	1
605	1
609	1
613	1
640	1
641	1
649	1
661	1
667	1
679	1
680	1
688	1
705	1
712	1
725	1
751	1
755	1
816	1
842	1
878	1
899	1
921	1
928	1
939	1
958	1
971	1
973	1
1035	1
1088	1
1154	1
1265	1
1292	1
1294	1
1372	1
1571	1
1642	1
1646	1
1863	1
1878	1
2146	1

APPENDIX A4 (CONTINUED)

WORD-PAIR FREQUENCY DISTIBUTIONS

1 BATCH = 1536 ABSTRACTS
F = FREQUENCY = NUMBER OF ABSTRACTS
P(F) = NUMBER OF WORD-PAIRS WITH FREQUENCY F
(OMITTING REPETITIONS WITHIN AN ABSTRACT)

BATCH 1

F	P(F)
0	443121
1	38778
2	9325
3	3461
4	1781
5	942
6	609
7	336
8	271
9	206
10	138
11	100
12	82
13	57
14	51
15	46
16	31
17	15
18	21
19	16
20	15
21	16
22	8
23	15
24	9
25	9
26	5
28	7
29	3
30	3
31	2
32	3
33	2
35	2
36	1
38	1
39	3
41	2
42	1
43	2
46	1
47	1
48	1
119	1

BATCHES 1-2

F	P(F)
0	410066
1	53356
2	16297
3	7082
4	3801
5	2376
6	1540
7	1057
8	768
9	555
10	428
11	339
12	256
13	216
14	164
15	157
16	139
17	104
18	103
19	79
20	65
21	62
22	49
23	38
24	35
25	36
26	27
27	22
28	20
29	30
30	16
31	15
32	13
33	19
34	12
35	12
36	15
37	3
38	7
39	13
40	9
41	3
42	7
43	7
44	5
45	6
46	4
47	7
48	2
49	4
50	4
51	2
52	4
54	3
55	4
56	3
58	1
59	2
60	1
61	1
62	3
63	3
64	1
65	2
66	1
67	1
68	1
71	3
72	1
74	1
77	1
81	1
82	1
83	1
84	3
85	1
90	1
100	1
108	1
213	1

BATCHES 1-4

F	P(F)	F	P(F)
0	355154	71	6
1	70954	72	7
2	27141	73	11
3	13663	74	3
4	8130	75	10
5	5291	76	3
6	3712	77	1
7	2654	78	9
8	2066	79	1
9	1581	80	4
10	1272	81	4
11	1039	82	4
12	820	83	3
13	706	84	4
14	596	85	1
15	482	86	3
16	430	87	1
17	321	88	6
18	307	89	8
19	294	91	5
20	277	92	2
21	213	93	2
22	214	95	2
23	191	96	2
24	133	97	5
25	139	98	2
26	121	99	2
27	108	100	3
28	90	101	1
29	101	102	3
30	95	103	1
31	73	104	2
32	69	105	1
33	64	106	2
34	46	107	2
35	62	108	1
36	53	109	2
37	47	110	3
38	35	111	3
39	36	112	1
40	43	115	1
41	45	116	1
42	36	117	2
43	26	118	1
44	27	119	1
45	29	122	1
46	23	124	1
47	25	126	1
48	23	129	1
49	22	131	1
50	17	132	1
51	15	133	1
52	20	135	1
53	21	138	1
54	13	142	2
55	20	144	1
56	11	145	1
57	11	149	1
58	13	159	1
59	10	165	1
60	16	166	1
61	11	168	1
62	7	175	2
63	14	183	1
64	9	190	1
65	8	191	1
66	7	192	1
67	7	201	1
68	7	207	1
69	8	219	1
70	6	381	1

BATCHES 1-8

F	P(F)	F	P(F)	F	P(F)	F	P(F)
0	296779	63	39	126	6	195	3
1	78712	64	50	127	4	196	2
2	36475	65	29	128	6	198	3
3	20741	66	37	129	3	199	1
4	13487	67	46	130	5	200	1
5	9396	68	41	131	5	201	4
6	6941	69	28	132	1	202	1
7	5197	70	29	133	2	203	1
8	3967	71	38	134	6	204	2
9	3313	72	28	135	3	206	2
10	2668	73	35	136	3	208	1
11	2231	74	30	137	7	210	2
12	1931	75	25	138	4	211	1
13	1613	76	27	139	1	213	1
14	1445	77	21	140	9	214	1
15	1252	78	18	141	3	215	1
16	1069	79	21	142	3	219	1
17	923	80	7	143	2	220	1
18	814	81	24	144	5	221	1
19	742	82	12	145	7	222	2
20	668	83	18	146	2	223	1
21	625	84	15	147	4	226	2
22	554	85	16	148	3	228	1
23	532	86	24	149	3	231	1
24	440	87	22	150	2	232	3
25	432	88	15	151	3	235	1
26	373	89	17	152	4	237	1
27	365	90	15	153	3	239	1
28	292	91	19	154	3	240	2
29	332	92	18	155	2	242	2
30	242	93	18	156	4	243	1
31	272	94	12	157	2	244	1
32	245	95	10	158	5	245	1
33	210	96	15	159	2	247	2
34	201	97	13	160	3	254	1
35	180	98	10	161	8	259	1
36	186	99	6	162	2	263	1
37	158	100	13	163	1	267	1
38	147	101	9	164	4	274	1
39	165	102	11	165	4	278	1
40	140	103	14	166	2	287	1
41	133	104	9	167	4	288	1
42	121	105	13	168	2	297	1
43	90	106	14	169	2	304	1
44	120	107	11	170	2	305	2
45	83	108	19	171	2	309	2
46	97	109	10	172	1	315	1
47	92	110	10	173	3	321	1
48	78	111	10	174	2	322	1
49	80	112	6	175	1	324	1
50	75	113	7	176	3	330	1
51	82	114	12	177	4	351	1
52	77	115	7	178	3	359	1
53	67	116	5	179	2	368	1
54	64	117	8	183	2	382	1
55	71	118	6	184	2	385	1
56	63	119	3	185	3	397	1
57	54	120	3	187	2	399	1
58	48	121	9	188	2	449	1
59	44	122	6	189	1	461	1
60	56	123	4	191	1	488	1
61	45	124	4	194	2	≥512	27
62	48	125	6				

APPENDIX A5

BINARY WORD CONNECTION MATRIX G3

0	0	ACCELERA
1	1	ACOUSTIC
	20	ASYMPTOT
	86	ELECTROM
	212	PERPENDI
	350	CONOUCT
	516	CYLIND
	818	WEDGE
	954	WAVE
2	2	ACTIVITY
	61	CORRELAT
	285	STATISTI
	467	SUNSPOT
	678	CYCLE
	960	YEAR
3	3	ADIA8ATI
	15	APPROXIM
	421	PASSAGE
	730	LEVEL
	737	MASER
4	4	ADMITTAN
	128	IMPEDANC
	248	REPRESEN
	298	TERMINAL
	412	NETWORK
	581	PASSIV
	739	MATRI
5	5	ADVANTAG
	59	CONVENTI
	201	OPERATIO
	639	TRIODE
	774	RELAY
	935	STAB
6	6	AFTERNOO
	178	MIONIGHT
	196	NONSEASO
	362	DAYTIME
	372	DIURNAL
	564	MEOIAN
	678	CYCLE
	728	LAYER
7	7	ALPHANUM
	39	CHARACTE
8	8	ALTERNAT
	796	SOLVE
9	9	AMPLITUD
10	10	ANALOGUE
	11	ANALYSER
	172	MATHEMAT
	511	COMPUT
11	10	ANALOGUE
	11	ANALYSER
	71	DIFFEREN
12	12	ANISOTRO
	70	DIELECTR
	86	ELECTROM
	121	HOMOGENE
	399	ISOTROP
	425	PERMEA8
	565	MEDIUM
	635	TENSOR
	703	FIELD
	741	MEDIA
	830	AXES
13	13	APPARATU
	608	SEARCH
	925	SCAN
14	14	APPARENT
	481	VIRTUAL
	983	LUN
15	3	ADIABATI
	15	APPROXIM
	112	GAUSSIAN
	454	SEGMENT
	700	EXPAN
16	16	ARGUMENT
	149	INTERPLA
17	17	ARTIFICI
	21	ATMOSPHE
	260	SATELLIT
	460	SPUTNIK
	693	EARTH
	755	OR8IT
18	18	ASSOCIAT
	205	OUT8URST
	597	RECORD
	880	HISS
19	19	ASYMMETR
	112	GAUSSIAN
	257	RESPONSE
	905	PASS
20	1	ACOUSTIC
	20	ASYMPTOT
	141	INTEGRAL
	271	SOLUTION
	513	CONVEX
	538	FINITE
	606	SCALAR
	643	UNIQUE
	699	EXACT
	700	EXPAN
	775	REVOL
21	17	ARTIFICI
	21	ATMOSPHE
	159	ISOTHERM
	163	LIGHTNIN
	236	RADIOSON
	416	ORBITAL
	580	OXYGEN
	672	CLOUD
	813	UPPER
	850	DRAG
	956	WIND
	963	AIR
22	22	ATTACHME
	42	COEFFICI
	87	ELECTRON
	106	FORMATIO
	159	ISOTHERM
	240	RECOMBIN
	524	DIFFUS
23	23	ATTENUAT
	520	DESIGN
	831	8AND
	867	FILT
	893	LOSS
	997	DB
24	24	AUTOMATI
	364	DECIMAL
	441	READING
24	632	SYSTEM
	854	DRUM
25	25	AVALANCH
26	26	8ACKGROU
	900	NOIS
27	27	8ACKWARD
	954	WAVE
28	28	8ALANCED
	60	CONVERSI
	497	8ALANC
	543	HEATER
	914	PULL
29	29	8ANDWIOT
	108	FREQUENC
	433	PRODUCT
	491	AMPLIF
	572	NARROW
	704	FIGUR
	754	OPTIM
	831	8AND
	874	GAIN
	915	PUMP
30	30	8IDIRECT
	356	CONTROL
	630	SWITCH
	977	GAT
31	31	8IFURCAT
	544	HEIGHT
	728	LAYER
32	32	8ISTABLE
	33	8LOCKING
	183	MONOSTA8
	190	MULTIVI8
	201	OPERATIO
	314	TRANSIST
	345	CIRCUIT
	477	TRIGGER
	630	SWITCH
	689	DIODE
	869	FLIP
33	32	8ISTABLE
	33	8LOCKING
	183	MONOSTA8
	190	MULTIVI8
	203	OSCILLAT
	326	WAVEFORM
	477	TRIGGER
	636	TIMING
	922	RISE
34	34	BREMSSTR
	216	PHOTOELE
	879	HIGH
	987	RAY
	999	X
35	35	CALCULAT
	380	EVALUAT
36	36	CALI8RAT
37	37	CENTIMET
	328	WAVELENG
	482	VISI8LE
38	38	CERENKOV
	70	DIELECTR
	235	RADIATIO

38	571	MOVING
39	7	ALPHANUM
	39	CHARACTE
	781	SELEN
	816	WAFER
40	40	CIRCULAR
	606	SCALAR
	647	VECTOR
	694	ELLIP
	931	SLIT
	967	BOD
41	41	CIRCULAT
42	22	ATTACHME
	42	COEFFICI
	240	RECOMBIN
	286	STOCHAST
	580	OXYGEN
43	43	COINCIDE
	742	MEMOR
44	44	COLLISIO
	87	ELECTRON
	121	HOMOGENE
	123	HYDRODYN
	246	RELATIVI
	316	TRANSPOR
	401	KINETIC
	406	MAGNETO
	413	NEUTRAL
	621	STATIC
	725	IONIC
	876	GYRO
45	45	COMMUNIC
	337	BEARING
46	46	COMPARIS
47	47	COMPATIB
	601	RESONA
48	48	COMPENSA
	101	FEEDBACK
	301	THERMIST
	461	STABILI
	491	AMPLIF
49	49	COMPLEME
	348	COLLECT
	869	FLIP
50	50	COMPONEN
	65	CRYOTRON
	308	TOLERANC
	816	WAFER
	825	ADHE
51	51	COMPRESS
52	52	CONCENTR
	191	NITROGEN
	240	RECOMBIN
	519	DENSIT
	580	OXYGEN
	979	ION
53	53	CONDENSE
	341	CAPACIT
54	54	CONDITIO
	194	NONPERIO
	339	BOUNDAR
55	55	CONFIGUR
	377	EMITTER
	109	FUNCTION
	673	CONIC
56	56	CONFORMA
56	312	TRANSFOR
57	57	CONJUNCT
58	58	CONSTITU
	191	NITROGEN
	829	ATOM
59	5	ADVANTAG
	59	CONVENTI
	101	FEEDBACK
	257	RESPONSE
	370	DISTORT
	378	EQUALIZ
	462	STAGGER
	546	HYBRID
	865	FEED
	874	GAIN
60	28	BALANCED
	60	CONVERSI
	83	EFFICIEN
	357	CONVERT
61	2	ACTIVITY
	61	CORRELAT
	285	STATISTI
62	62	CORRESPO
63	63	CORRUGAT
	73	DIMENSIO
	259	ROUGHNES
	270	SINUSOID
	278	SPHEROID
	452	SCATTER
	469	SURFACE
	703	FIELD
	786	SHEET
	954	WAVE
64	64	CRITICAL
	740	MAXIM
65	50	COMPONEN
	65	CRYOTRON
	252	RESISTOR
	291	SUPERCON
	365	DEPOSIT
	742	MEMOR
	823	WOUND
	866	FILM
	938	STOR
66	66	CYCLOTRO
	98	EXTERNAL
	131	INCOHERE
	254	RESONANC
	587	PLASMA
	621	STATIC
67	67	DEMODULA
	270	SINUSOID
	370	DISTORT
	407	MODULAT
68	68	DEPARTUR
69	69	DERIVATI
	109	FUNCTION
	248	REPRESEN
70	12	ANISOTRO
	38	CERENKOV
	70	DIELECTR
	88	ELECTROS
	399	ISOTROP
	425	PERMEAB
	565	MEDIUM
	635	TENSOR
	741	MEDIA
71	11	ANALYSER
	71	DIFFEREN
72	72	DIFFRACT
	86	ELECTROM
	271	SOLUTION
	339	BOUNDAR
	395	INFINIT
	513	CONVEX
	606	SCALAR
	607	SCREEN
	700	EXPAN
	759	PLANE
	818	WEDGE
	931	SLIT
	954	WAVE
	967	BOD
73	63	CORRUGAT
	73	DIMENSIO
	759	PLANE
74	74	DIRECTIO
75	75	DIRECTIV
	490	AERIAL
	658	ANALY
76	76	DISCHARG
	163	LIGHTNIN
	227	PRESSURE
	387	GASEOUS
	566	MERCUR
	765	PROBE
	799	SPARK
77	77	DISCONTI
78	78	DISCRETE
	111	GALACTIC
	573	NEBULA
	615	SOURCE
79	79	DISSIPAT
80	80	DISTRIBU
	113	GEOGRAPH
	159	ISOTHERM
	459	SPATIAL
	500	BRIGHT
81	81	ECCENTRI
	416	ORBITAL
	755	ORBIT
	775	REVOL
82	82	EFFECTIV
83	60	CONVERSI
	83	EFFICIEN
	579	OUTPUT
	763	POWER
	951	VOLT
84	84	ELECTRIC
	316	TRANSPOR
	746	MOBIL
	985	NET
85	85	ELECTROD
	88	ELECTROS
	830	AXES
86	1	ACOUSTIC
	12	ANISOTRO
	72	DIFFRACT
	86	ELECTROM
	277	SPHERICA
	278	SPHEROID
	395	INFINIT
	399	ISOTROP
	565	MEDIUM
	607	SCREEN
	617	SPHERE
	741	MEDIA
	818	WEDGE

86	954	WAVE
87	22	ATTACHME
	44	COLLISIO
	87	ELECTRON
	191	NITROGEN
	262	SECONOAR
	316	TRANSPOR
	413	NEUTRAL
	434	PROFILE
	524	DIFFUS
	634	TARGET
	644	VALENC
	710	GASES
	765	PROBE
	829	ATOM
	979	ION
88	70	DIELECTR
	85	ELECTROD
	88	ELECTROS
	225	POTENTIA
	807	THEOR
	830	AXES
89	89	ENGINEER
	404	MACHINE
	435	PROGRAM
	511	COMPUT
90	90	EQUILIBR
91	91	EQUIVALE
92	92	EVAPORAT
	743	METAL
	866	FILM
	903	OXIO
	948	VACU
93	93	EXCHANGE
	144	INTERACT
	532	ENERGY
94	94	EXCITATI
	381	EXCITED
	644	VALENC
95	95	EXOSPHER
	651	WHISTL
	693	EARTH
96	96	EXPONENT
97	97	EXTENSIO
98	66	CYCLOTRO
	98	EXTERNAL
	332	ANNULAR
99	99	EXTINCTI
	391	INCIOEN
	430	POLARIZ
	444	REFLECT
	452	SCATTER
	576	OBLIQU
	577	ORIENT
	694	ELLIP
	954	WAVE
100	100	EXTRAORO
	243	REFLEXIO
	406	MAGNETO
	645	VALLEY
	808	TRACE
	876	GYRO
101	48	COMPENSA
	59	CONVENTI
	101	FEEOBACK
	293	SYNCHRON
	461	STABILI
	499	BRIOGE
	782	SERVO
101	852	DRIV
	892	LOOP
102	102	FERROELE
	341	CAPACIT
103	103	FERROMAG
	166	MAGNETIC
	167	MAGNE,IS
	193	NONMAGNE
	405	MAGNETI
	807	THEOR
	934	SPIN
104	104	FILAMENT
	361	CURRENT
105	105	FLUCTUAT
	927	SHOT
106	22	ATTACHME
	106	FORMATIO
	153	IONIZATI
	240	RECOMBIN
	580	OXYGEN
	826	ALTI
	933	SOFT
107	107	FRACTION
108	29	BANDWIDT
	108	FREQUENC
	263	SELECTIV
	478	TUNABLE
	488	ADJUST
	572	NARROW
	601	RESONA
	806	SWEEP
	890	LOCK
	943	TONE
	978	IOL
	984	MIX
109	56	CONFORMA
	69	DERIVATI
	109	FUNCTION
	172	MATHEMAT
	199	NUMERICA
	221	POLYNOMI
	311	TRANSFER
	313	TRANSIEN
	380	EVALUAT
	911	POLE
	920	REAL
110	110	FUNOAMEN
	386	FOURIER
	680	OAMPI
111	78	OISCRETE
	111	GALACTIC
	235	RAOIATIO
	573	NEBULA
	615	SOURCE
	709	GALAX
	771	RAOIO
	871	FLUX
	964	ARC
112	15	APPROXIM
	19	ASYMMETR
	112	GAUSSIAN
	984	MIX
113	80	DISTRIBU
	113	GEOGRAPH
	120	HEMISPHE
	162	LATITUDE
	463	STATION
	679	OAILY
	797	SOUTH
	822	WORLD
114	114	GEOMAGNE
114	176	MICROPUL
	371	DISTURB
	379	EQUATOR
	529	OYNAMO
	591	PULSAT
	712	GIANT
	733	LOCAL
	762	POLAR
	797	SOUTH
	803	STORM
	894	MAIN
	965	BAY
115	115	GEOPHYSI
	148	INTERNAT
	857	EAST
	960	YEAR
116	116	GERMANIU
	160	JUNCTION
	599	RECTIF
	792	SILIC
117	117	GRAOIENT
	159	ISOTHERM
	325	VERTICAL
	529	OYNAMO
118	118	HAMILTON
119	119	HARMONIC
120	113	GEOGRAPH
	120	HEMISPHE
	162	LATITUOE
	324	VARIATIO
	463	STATION
	752	NORTH
	797	SOUTH
121	12	ANISOTRO
	44	COLLISIO
	121	HOMOGENE
	401	KINETIC
	621	STATIC
	635	TENSOR
	703	FIELO
	707	FLUIO
	967	BOD
122	122	HORIZONT
	185	MOVEMENT
	325	VERTICAL
	956	WINO
123	44	COLLISIO
	123	HYDROOYN
	387	GASEOUS
	406	MAGNETO
	587	PLASMA
	725	IONIC
	726	IONIZ
124	124	HYDROGEN
	765	PROBE
125	125	HYDROMAG
	591	PULSAT
	707	FLUID
126	126	HYPERBOL
127	127	HYPOTHES
128	4	AOMITTAN
	128	IMPEDANC
	138	INSERTIO
	233	QUADRIPO
	298	TERMINAL
	487	ACTIVE
	498	BRANCH
	502	CASCAO
	556	LADOER
	719	IMAGE

128	738	MATCH
129	129	IMPERFEC
	469	SURFACE
	699	EXACT
130	130	IMPURITY
	644	VALENC
131	66	CYCLOTRO
	131	INCOHERE
	246	RELATIVI
	413	NEUTRAL
	452	SCATTER
	621	STATIC
132	132	INCREMEN
133	133	INDEPENO
134	134	INDUCTAN
	135	INDUCTOR
	341	CAPACIT
	345	CIRCUIT
	393	INDUCTI
	823	WOUNO
	841	COIL
135	134	INOUCTAN
	135	INDUCTOR
	341	CAPACIT
	439	REACTIV
136	136	INFORMAT
	441	REAOING
	484	ACCESS
	511	COMPUT
	632	SYSTEM
	687	OIGIT
	742	MEMOR
	800	SPEEO
	938	STOR
	940	TAPE
137	137	INHOMOGE
138	128	IMPEOANC
	138	INSERTIO
	233	QUAORIPO
	299	TERMINAT
	498	BRANCH
	556	LAOOER
	719	IMAGE
	831	BANO
	867	FILT
	893	LOSS
	905	PASS
139	139	INSTABIL
140	140	INSTRUME
	605	ROCKET
141	20	ASYMPTOT
	141	INTEGRAL
	271	SOLUTION
	696	EQUAT
	700	EXPAN
	796	SOLVE
	818	WEOGE
142	142	INTEGRAT
143	143	INTENSIT
	149	INTERPLA
	300	TERRESTR
	515	COSMIC
144	93	EXCHANGE
	144	INTERACT
	167	MAGNETIS
	382	EXCITON
	532	ENERGY
	569	MOMENT

145	145	INTERFER
	235	RAOIATIO
	295	TELESCOP
	328	WAVELENG
	490	AERIAL
	523	OIAMET
	615	SOURCE
	661	ASTRO
	771	RAOIO
	993	SUN
146	146	INTERMED
147	147	INTERNAL
	377	EMITTER
	865	FEEO
148	115	GEOPHYSI
	148	INTERNAT
	960	YEAR
149	16	ARGUMENT
	143	INTENSIT
	149	INTERPLA
	300	TERRESTR
	358	CORPUSC
	515	COSMIC
	693	EARTH
	798	SPACE
	828	AREA
	987	RAY
150	150	INTERVAL
	597	RECORO
	811	TRAIN
	942	TIME
151	151	INVARIAN
152	152	INVERSIO
	222	POPULATI
	801	SPLIT
153	106	FORMATIO
	153	IONIZATI
	154	IONOGRAM
	155	IONOSONO
	185	MOVEMENT
	280	SPORAOIC
	379	EQUATOR
	524	OIFFUS
154	153	IONIZATI
	154	IONOGRAM
	155	IONOSONO
	196	NONSEASO
	280	SPORAOIC
	325	VERTICAL
	645	VALLEY
155	153	IONIZATI
	154	IONOGRAM
	155	IONOSONO
	272	SOUNOING
	597	RECORO
	619	SPREAO
156	156	IONOSPHE
	185	MOVEMENT
	243	REFLEXIO
	272	SOUNDING
	325	VERTICAL
	362	OAYTIME
	372	OIURNAL
	379	EQUATOR
	463	STATION
	481	VIRTUAL
	493	ARCTIC
	529	OYNAMO
	609	SEASON
	619	SPREAO
	627	SUMMER
	652	WINTER

156	679	OAILY
	752	NORTH
	797	SOUTH
157	157	IRREGULA
	243	REFLEXIO
	319	TURBULEN
	494	ASPECT
	537	FADING
	619	SPREAO
	655	ALIGN
	858	ECHO
158	158	IRREVERS
	303	THERMODY
	832	BASE
	901	NOTE
159	21	ATMOSPHE
	22	ATTACHME
	80	OISTRIBU
	117	GRAOIENT
	159	ISOTHERM
	296	TEMPERAT
	390	HEATING
	780	SCALE
160	116	GERMANIU
	160	JUNCTION
	314	TRANSIST
	546	HYBRIO
	636	ALLOY
	792	SILIC
161	161	LABORATO
162	113	GEOGRAPH
	120	HEMISPHE
	162	LATITUOE
	184	MORPHOLO
	379	EQUATOR
	679	OAILY
	762	POLAR
	768	QUIET
	822	WORLO
	960	YEAR
163	21	ATMOSPHE
	76	OISCHARG
	163	LIGHTNIN
	473	THUNOER
	603	RETURN
	626	STROKE
	672	CLOUO
	706	FLASH
164	164	LONGITUD
	474	TORSION
	587	PLASMA
165	165	LUMINESC
	375	ELECTRO
166	103	FERROMAG
	166	MAGNETIC
	167	MAGNETIS
	385	FERRITE
	405	MAGNETI
	451	SATURAT
	569	MOMENT
	707	FLUIO
	708	FORCE
	711	GAUSS
	854	ORUM
	877	HEAD
167	103	FERROMAG
	144	INTERACT
	166	MAGNETIC
	167	MAGNETIS
	569	MOMENT
	934	SPIN
168	168	MAGNETOM

168	563	MEASUR
169	169	MAGNETOR
	264	SEMICOND
	350	CONDUCT
	703	FIELD
170	170	MAGNETOS
	474	TORSION
171	171	MAGNITUD
172	10	ANALOGUE
	109	FUNCTION
	172	MATHEMAT
	214	PHOTOCEL
	248	REPRESEN
	538	FINITE
	622	STEADY
	631	SYMBOL
	647	VECTOR
	696	EQUAT
173	173	MECHANIC
174	174	MECHANIS
175	175	MICROMIN
	181	MINIATUR
	345	CIRCUIT
	520	DESIGN
	735	LOGIC
	748	MODUL
176	114	GEOMAGNE
	176	MICROPUL
	591	PULSAT
177	177	MICROWAV
	276	SPECTROS
178	6	AFTERNOO
	178	MIDNIGHT
	196	NONSEASO
	582	PERIOD
	733	LOCAL
	740	MAXIM
	965	BAY
179	179	MILLIMET
	289	SUBMILLI
	388	GENERAT
	954	WAVE
180	180	MILLIMIC
	388	GENERAT
	766	PULSE
	922	RISE
181	175	MICROMIN
	181	MINIATUR
182	182	MOMENTUM
	532	ENERGY
183	32	BISTABLE
	33	BLOCKING
	183	MONOSTAB
	190	MULTIVIB
	314	TRANSIST
	345	CIRCUIT
	477	TRIGGER
	610	SELECT
	835	BIAS
	869	FLIP
184	162	LATITUDE
	184	MORPHOLO
	371	DISTURB
	489	ADVANC
	803	STORM
	994	TID
[illegible]	122	HORIZONT

185	153	IONIZATI
	156	IONOSPHE
	185	MOVEMENT
186	186	MULTILAY
187	187	MULTIPLI
188	188	MULTIPOL
	739	MATRI
189	189	MULTISTA
	257	RESPONSE
	417	OVERSHO
	462	STAGGER
190	32	BISTABLE
	33	BLOCKING
	183	MONOSTAB
	190	MULTIVIB
	345	CIRCUIT
	477	TRIGGER
	869	FLIP
191	52	CONCENTR
	58	CONSTITU
	87	ELECTRON
	191	NITROGEN
	238	REACTION
	495	ATOMIC
	545	HELIUM
	580	OXYGEN
	710	GASES
	973	ION
192	192	NONLINEA
	990	ROW
193	103	FERROMAG
	193	NONMAGNE
	934	SPIN
194	54	CONDITIO
	194	NONPERIO
195	195	NONRECIP
	376	ELEMENT
	581	PASSIV
196	6	AFTERNOO
	154	IONOGRAM
	178	MIDNIGHT
	196	NONSEASO
	324	VARIATIO
	466	SUNRISE
	492	ANOMAL
	519	DENSIT
	544	HEIGHT
	609	SEASON
	679	DAILY
	728	LAYER
197	197	NONUNIFO
198	198	NORMALIZ
	299	TERMINAT
	633	TABULA
	867	FILT
199	109	FUNCTION
	199	NUMERICA
200	200	OBSERVAT
	295	TELESCOP
	300	TERRESTR
	366	DETONAT
	400	JUPITER
	500	BRIGHT
	573	NEBULA
	649	VISUAL
	661	ASTRO
	736	MAJOR
	768	QUIET
	797	SOUTH

201	5	ADVANTAG
	32	BISTABLE
	201	OPERATIO
	292	SUPERREG
	663	BASIC
	735	LOGIC
	869	FLIP
	962	ADD
202	202	ORTHOGON
203	33	BLOCKING
	203	OSCILLAT
	381	EXCITED
	410	NATURAL
	601	RESONA
	680	DAMPI
	890	LOCK
	926	SELF
204	204	OSCILLOG
	535	EXTREM
205	18	ASSOCIAT
	205	OUTBURST
	667	BURST
	698	ERUPT
	705	FLARE
	794	SOLAR
	803	STORM
206	206	PARALLEL
207	207	PARAMAGN
	254	RESONANC
208	208	PARTICLE
	246	RELATIVI
	309	TRAJECTO
	571	MOVING
	590	PROTON
	669	CHARG
	708	FORCE
	757	OUTER
	834	BELT
	944	TRAP
209	209	PENETRAT
	685	DEPTH
210	210	PERFORMA
211	211	PERMANEN
212	1	ACOUSTIC
	212	PERPENDI
	647	VECTOR
	759	PLANE
	786	SHEET
213	213	PHENOMEN
	883	JUMP
214	172	MATHEMAT
	214	PHOTOCEL
	302	THERMOCO
215	215	PHOTOCON
	375	ELECTRO
216	34	BREMSSTR
	216	PHOTOELE
	743	METAL
	909	PHOT
217	217	PHOTOGRA
	596	RECEIV
	597	RECORD
	649	VISUAL
218	218	PHOTOSEN
	456	SHAPING
	675	COUNT

218	753	OPTIC
219	219	PIEZOELE
	360	CRYSTAL
220	220	POLARITY
221	109	FUNCTION
	221	POLYNOMI
	920	REAL
222	152	INVERSIO
	222	POPULATI
	737	MASER
	915	PUMP
223	223	POSITION
	504	CENTRE
224	224	POSITIVE
	411	NEGATIV
225	88	ELECTROS
	225	POTENTIA
	336	BARRIER
226	226	POTENTIO
	250	RESISTAN
227	76	DISCHARG
	227	PRESSURE
	387	GASEOUS
	424	PERIGEE
	545	HELIUM
	710	GASES
	765	PROBE
228	228	PRINCIPA
229	229	PROBABIL
230	230	PROGRESS
231	231	PROPAGAT
	406	MAGNETO
	741	MEDIA
	856	DUCT
	954	WAVE
232	232	PROPORTI
	398	INVERSE
233	128	IMPEDANC
	138	INSERTIO
	233	QUADRIPO
	298	TERMINAL
	340	CANONIC
	412	NETWORK
	498	BRANCH
	502	CASCAD
	581	PASSIV
	719	IMAGE
234	234	QUADRUPO
	415	NUCLEAR
	737	MASER
235	38	CERENKOV
	111	GALACTIC
	145	INTERFER
	235	RADIATIO
	328	WAVELENG
	363	DECIMET
	400	JUPITER
	500	BRIGHT
	514	CORONA
	573	NEBULA
	709	GALAX
	834	BELT
	871	FLUX
	988	RED
	993	SUN
236	21	ATMOSPHE
236	236	RADIOSON
	445	REFRACT
	563	MEASUR
	721	INDEX
	848	DATA
237	237	REACTANC
	439	REACTIV
	990	ROW
238	191	NITROGEN
	238	REACTION
	580	OXYGEN
239	239	RECIPROC
	298	TERMINAL
	581	PASSIV
	695	EQUAL
	904	PAIR
240	22	ATTACHME
	42	COEFFICI
	52	CONCENTR
	106	FORMATIO
	240	RECOMBIN
	495	ATOMIC
	580	OXYGEN
	863	FALL
241	241	REFERENC
242	242	REFLECTO
	295	TELESCOP
	418	PARABOL
243	100	EXTRAORD
	156	IONOSPHE
	157	IRREGULA
	243	REFLEXIO
244	244	REGENERA
	293	SYNCHRON
	849	DIVI
245	245	RELATIVE
246	44	COLLISIO
	131	INCOHERE
	208	PARTICLE
	246	RELATIVI
	401	KINETIC
	542	GROWTH
	551	INJECT
	571	MOVING
	669	CHARG
	708	FORCE
247	247	RELAXATI
	254	RESONANC
248	4	ADMITTAN
	69	DERIVATI
	172	MATHEMAT
	248	REPRESEN
	349	COMPLEX
	739	MATRI
249	249	RESIDUAL
250	226	POTENTIO
	250	RESISTAN
251	251	RESISTIV
	252	RESISTOR
	341	CAPACIT
	439	REACTIV
252	65	CRYOTRON
	251	RESISTIV
	252	RESISTOR
	436	PROTECT
	501	CARBON
	823	WOUND
	973	FIX
253	253	RESOLUTI
	490	AERIAL
	600	RESOLV
	879	HIGH
	925	SCAN
254	66	CYCLOTRO
	207	PARAMAGN
	247	RELAXATI
	254	RESONANC
	286	STOCHAST
	415	NUCLEAR
	702	FERRI
255	255	RESONATO
	410	NATURAL
	668	CAVIT
256	19	ASYMMETR
	59	CONVENTI
	189	MULTISTA
	256	RESPONSE
	263	SELECTIV
	313	TRANSIEN
	378	EQUALIZ
	417	OVERSHO
	454	SEGMENT
	462	STAGGER
	471	TELEVIS
	782	SERVO
	883	JUMP
	910	PLOT
	937	STEP
257	257	RESTRICT
	417	OVERSHO
	990	ROW
258	258	REVERSIB
	665	BINAR
	675	COUNT
259	63	CORRUGAT
	259	ROUGHNES
	444	REFLECT
	469	SURFACE
	563	MEASUR
260	17	ARTIFICI
	260	SATELLIT
	366	DETONAT
	384	FARADAY
	416	ORBITAL
	424	PERIGEE
	460	SPUTNIK
	557	LAUNCH
	693	EARTH
	755	ORBIT
	815	VEHIC
	850	DRAG
261	261	SCINTILL
	936	STAR
262	87	ELECTRON
	262	SECONDAR
	338	BOMBARD
	432	PRIMARY
	531	EMISSI
	532	ENERGY
	634	TARGET
	743	METAL
	903	OXID
263	108	FREQUENC
	257	RESPONSE
	263	SELECTIV
	378	EQUALIZ
	462	STAGGER
	499	BRIDGE
	520	DESIGN
	572	NARROW
	905	PASS

264	169	MAGNETOR
	264	SEMICOND
	342	CARRIER
265	265	SEMIDIUR
	324	VARIATIO
	529	DYNAMO
	908	PHAS
	983	LUN
	994	TID
266	266	SENSITIV
	494	ASPECT
267	267	SEQUENCE
268	268	SEQUENTI
269	269	SIMULTAN
	496	AURORA
	858	ECHO
270	63	CORRUGAT
	67	DEMODULA
	270	SINUSOID
	370	DISTORT
271	20	ASYMPTOT
	72	DIFFRACT
	141	INTEGRAL
	271	SOLUTION
	278	SPHEROID
	339	BOUNDAR
	395	INFINIT
	538	FINITE
	606	SCALAR
	607	SCREEN
	643	UNIQUE
	696	EQUAT
	699	EXACT
	700	EXPAN
	796	SOLVE
272	155	IONOSOND
	156	IONOSPHE
	272	SOUNDING
	325	VERTICAL
	481	VIRTUAL
	858	ECHO
273	273	SPECIMEN
274	274	SPECTROG
	276	SPECTROS
275	275	SPECTROM
	276	SPECTROS
276	177	MICROWAV
	274	SPECTROG
	275	SPECTROM
	276	SPECTROS
277	86	ELECTROM
	277	SPHERICA
	350	CONDUCT
	516	CYLIND
	759	PLANE
278	63	CORRUGAT
	86	ELECTROM
	271	SOLUTION
	278	SPHEROID
	606	SCALAR
	700	EXPAN
	967	BOD
279	279	SPONTANE
	506	COHERE
	531	EMISSI
280	153	IONIZATI
	154	IONOGRAM
280	280	SPORADIC
	325	VERTICAL
	544	HEIGHT
	597	RECORD
	858	ECHO
281	281	SPURIOUS
	520	DESIGN
	610	SELECT
	898	MODE
282	282	STANDARD
283	283	STANDING
	381	EXCITED
	587	PLASMA
	954	WAVE
284	284	STATIONA
	538	FINITE
	542	GROWTH
	696	EQUAT
285	2	ACTIVITY
	61	CORRELAT
	285	STATISTI
	467	SUNSPOT
	582	PERIOD
	803	STORM
286	42	COEFFICI
	254	RESONANC
	286	STOCHAST
	595	RANDOM
	616	SPECTR
	807	THEOR
	934	SPIN
287	287	STRENGTH
	703	FIELD
288	288	STRUCTUR
	868	FINE
289	179	MILLIMET
	289	SUBMILLI
	388	GENERAT
	833	BEAM
	954	WAVE
290	290	SUCCESSI
	626	STROKE
291	65	CRYOTRON
	291	SUPERCON
	365	DEPOSIT
292	201	OPERATIO
	292	SUPERREG
	419	PARAMET
	491	AMPLIF
	689	DIODE
	900	NOIS
293	101	FEEDBACK
	244	REGENERA
	293	SYNCHRON
	370	DISTORT
	403	LEADING
	724	INPUT
	943	TONE
294	294	SYNTHESI
	340	CANONIC
	376	ELEMENT
	412	NETWORK
295	145	INTERFER
	200	OBSERVAT
	242	REFLECTO
	295	TELESCOP
	490	AERIAL
	523	DIAMET
	568	MIRROR
295	596	RECEIV
	661	ASTRO
	771	RADIO
	925	SCAN
296	159	ISOTHERM
	296	TEMPERAT
	304	THERMOEL
	329	AMBIENT
	924	ROOM
297	297	TEMPORAL
	372	DIURNAL
	609	SEASON
298	4	ADMITTAN
	128	IMPEDANC
	233	QUADRIPO
	239	RECIPROC
	298	TERMINAL
	412	NETWORK
	581	PASSIV
299	138	INSERTIO
	198	NORMALIZ
	299	TERMINAT
	439	REACTIV
	498	BRANCH
	556	LADDER
	738	MATCH
	905	PASS
300	143	INTENSIT
	149	INTERPLA
	200	OBSERVAT
	300	TERRESTR
	515	COSMIC
	693	EARTH
	803	STORM
	987	RAY
	993	SUN
301	48	COMPENSA
	301	THERMIST
	329	AMBIENT
	461	STABILI
302	214	PHOTOCEL
	302	THERMOCO
303	158	IRREVERS
	303	THERMODY
304	296	TEMPERAT
	304	THERMOEL
305	305	THICKNES
	785	SHEAR
306	306	THRESHOL
307	307	THYRATRO
	345	CIRCUIT
	446	REGULAT
	450	SATURAB
	468	SUPPLIE
	579	OUTPUT
	951	VOLT
308	50	COMPONEN
	308	TOLERANC
	520	DESIGN
	950	VIEW
309	208	PARTICLE
	309	TRAJECTO
	669	CHARG
310	310	TRANSDUC
311	109	FUNCTION
	311	TRANSFER
	412	NETWORK
	911	POLE

312	56	CONFORMA
	312	TRANSFOR
	340	CANONIC
	386	FOURIER
	801	SPLIT
313	109	FUNCTION
	257	RESPONSE
	313	TRANSIEN
	417	OVERSHO
	559	LINEAR
	724	INPUT
	937	STEP
314	32	BISTABLE
	160	JUNCTION
	183	MONOSTAB
	314	TRANSIST
	345	CIRCUIT
	348	COLLECT
	377	EMITTER
	477	TRIGGER
	663	BASIC
	671	CLASS
	835	BIAS
	838	CHOP
	869	FLIP
	935	STAB
315	315	TRANSITI
316	44	COLLISIO
	84	ELECTRIC
	87	ELECTRON
	316	TRANSPOR
	401	KINETIC
	621	STATIC
	703	FIELD
	726	IONIZ
	746	MOBIL
	976	GAS
317	317	TRANSVER
	621	STATIC
	703	FIELD
318	318	TRAVELLI
	954	WAVE
319	157	IRREGULA
	319	TURBULEN
	567	METEOR
	780	SCALE
	810	TRAIL
320	320	UNIDIREC
321	321	UNIVERSA
322	322	UNSTABLE
323	323	VARIABLE
	973	FIX
324	120	HEMISPHE
	196	NONSEASO
	265	SEMIDIUR
	324	VARIATIO
	372	DIURNAL
	609	SEASON
	627	SUMMER
	652	WINTER
	679	DAILY
	882	HOUR
	960	YEAR
	994	TID
325	117	GRADIENT
	122	HORIZONT
	154	IONOGRAM
	156	IONOSPHE
	272	SOUNDING
	280	SPORADIC
325	325	VERTICAL
	544	HEIGHT
	728	LAYER
	780	SCALE
	983	LUN
326	33	BLOCKING
	326	WAVEFORM
	610	SELECT
	766	PULSE
327	327	WAVEGUID
328	37	CENTIMET
	145	INTERFER
	235	RADIATIO
	328	WAVELENG
	363	DECIMET
	744	METRE
329	296	TEMPERAT
	301	THERMIST
	329	AMBIENT
	403	LEADING
	844	COOL
330	330	ANALOGY
331	331	ANGULAR
	771	RADIO
332	98	EXTERNAL
	332	ANNULAR
	530	ECLIPS
333	333	ANTENNA
	490	AERIAL
	596	RECEIV
334	334	AVERAGE
335	335	BALLOON
	515	COSMIC
	826	ALTI
	987	RAY
336	225	POTENTIA
	336	BARRIER
	469	SURFACE
337	45	COMMUNIC
	337	BEARING
	384	FARADAY
338	262	SECONDAR
	338	BOMBARD
	531	EMISSI
339	54	CONDITIO
	72	DIFFRACT
	271	SOLUTION
	339	BOUNDAR
	538	FINITE
	643	UNIQUE
	673	CONIC
	759	PLANE
340	233	QUADRIPO
	294	SYNTHESI
	312	TRANSFOR
	340	CANONIC
	412	NETWORK
	581	PASSIV
	739	MATRI
341	53	CONDENSE
	102	FERROELE
	134	INDUCTAN
	135	INDUCTOR
	251	RESISTIV
	341	CAPACIT
	484	ACCESS
	791	SHUNT
	973	FIX
342	264	SEMICOND
	342	CARRIER
343	343	CENTRAL
	421	PASSAGE
	467	SUNSPOT
	582	PERIOD
344	344	CHANNEL
345	32	BISTABLE
	134	INDUCTAN
	175	MICROMIN
	183	MONOSTAB
	190	MULTIVIB
	307	THYRATRO
	314	TRANSIST
	345	CIRCUIT
	440	REACTOR
	456	SHAPING
	477	TRIGGER
	497	BALANC
	852	DRIV
346	346	CLASSIC
	635	TENSOR
	703	FIELD
	796	SOLVE
	807	THEOR
	830	AXES
347	347	CLASSIF
348	49	COMPLEME
	314	TRANSIST
	348	COLLECT
	377	EMITTER
349	248	REPRESEN
	349	COMPLEX
350	1	ACOUSTIC
	169	MAGNETOR
	277	SPHERICA
	350	CONDUCT
	513	CONVEX
	516	CYLIND
	607	SCREEN
	617	SPHERE
	644	VALENC
	699	EXACT
	707	FLUID
	786	SHEET
	930	SKIN
351	351	CONSERV
	532	ENERGY
352	352	CONSTAN
353	353	CONTACT
	761	POINT
354	354	CONTINU
355	355	CONTOUR
356	30	BIDIRECT
	356	CONTROL
	455	SENSING
	588	PRECIS
	589	PRESET
	750	MOTOR
357	60	CONVERSI
	357	CONVERT
	403	LEADING
	642	TUNNEL
	724	INPUT
358	149	INTERPLA
	358	CORPUSC
	705	FLARE

358	736	MAJDR
359	359	CORRECT
360	219	PIEZOELE
	360	CRYSTAL
	382	EXCITDN
	648	VIBRAT
	785	SHEAR
361	104	FILAMENT
	361	CURRENT
	393	INDUCTI
	603	RETURN
	722	INDUC
	971	EDD
362	6	AFTERNOO
	156	IONOSPHE
	362	DAYTIME
	372	DIURNAL
	463	STATION
	609	SEASON
	627	SUMMER
	652	WINTER
	751	NIGHT
363	24	AUTOMATI
	363	DECIMAL
	441	READING
	665	BINAR
	687	DIGIT
	940	TAPE
	962	ADD
	966	BIT
	968	COD
364	235	RADIATIO
	328	WAVELENG
	364	DECIMET
365	65	CRYOTRON
	291	SUPERCON
	365	DEPOSIT
	866	FILM
366	200	OBSERVAT
	260	SATELLIT
	366	DETONAT
	383	EXPLOSI
	415	NUCLEAR
	734	LOCAT
	944	TRAP
367	367	DEVELDP
368	368	DISPERS
369	369	DISPLAC
370	59	CONVENTI
	67	DEMODULA
	270	SINUSOID
	293	SYNCHRON
	370	DISTORT
	671	CLASS
371	114	GEOMAGNE
	184	MORPHOLO
	371	DISTURB
	712	GIANT
	733	LOCAL
	762	POLAR
372	6	AFTERNOO
	156	IONOSPHE
	297	TEMPORAL
	324	VARIATIO
	362	DAYTIME
	372	DIURNAL
	609	SEASON
	679	DAILY
	733	LOCAL
	768	QUIET
	797	SOUTH
	882	HOUR
373	373	DOPPLER
	384	FARADAY
	460	SPUTNIK
	788	SHIFT
374	374	DYNAMIC
375	165	LUMINESC
	215	PHOTOCON
	375	ELECTRO
376	195	NONRECIP
	294	SYNTHESI
	376	ELEMENT
377	55	CONFIGUR
	147	INTERNAL
	314	TRANSIST
	348	COLLECT
	377	EMITTER
	491	AMPLIF
	671	CLASS
	791	SHUNT
	835	BIAS
378	59	CONVENTI
	257	RESPONSE
	263	SELECTIV
	378	EQUALIZ
	471	TELEVIS
	520	DESIGN
379	114	GEOMAGNE
	153	IONIZATI
	156	IONOSPHE
	162	LATITUDE
	379	EQUATOR
	728	LAYER
	768	QUIET
380	35	CALCULAT
	109	FUNCTION
	380	EVALUAT
381	94	EXCITATI
	203	OSCILLAT
	283	STANDING
	381	EXCITED
	587	PLASMA
	668	CAVIT
382	144	INTERACT
	360	CRYSTAL
	382	EXCITON
	402	LATTICE
	648	VIBRAT
	807	THEOR
383	366	DETONAT
	383	EXPLOSI
	415	NUCLEAR
	553	ISLAND
384	260	SATELLIT
	337	BEARING
	373	DOPPLER
	384	FARADAY
	444	REFLECT
	519	DENSIT
	537	FADING
	777	ROTAT
385	166	MAGNETIC
	385	FERRITE
	405	MAGNETI
	845	CORE
386	110	FUNDAMEN
	312	TRANSFOR
	386	FOURIER
	622	STEADY
	643	UNIQUE
387	76	DISCHARG
387	123	HYDRODYN
	227	PRESSURE
	387	GASEOUS
	545	HELIUM
	587	PLASMA
	765	PROBE
388	179	MILLIMET
	180	MILLIMIC
	289	SUBMILLI
	388	GENERAT
	849	DIVI
	922	RISE
389	389	GEOMETR
390	159	ISOTHERM
	390	HEATING
391	99	EXTINCTI
	391	INCIDEN
	444	REFLECT
	576	OBLIQU
	931	SLIT
392	392	INDICAT
393	134	INDUCTAN
	361	CURRENT
	393	INDUCTI
394	394	INERTIA
	621	STATIC
395	72	DIFFRACT
	86	ELECTROM
	271	SOLUTION
	395	INFINIT
	513	CONVEX
	606	SCALAR
	607	SCREEN
	699	EXACT
	759	PLANE
	786	SHEET
	967	BOD
396	396	INITIAL
397	397	INSULAT
	743	METAL
398	232	PROPORTI
	398	INVERSE
399	12	ANISOTRO
	70	DIELECTR
	86	ELECTROM
	399	ISOTROP
	635	TENSDR
400	200	OBSERVAT
	235	RADIATIO
	400	JUPITER
	586	PLANET
	771	RADIO
401	44	COLLI
	121	HOMOGENE
	246	RELATIVI
	316	TRANSPOR
	401	KINETIC
	532	ENERGY
	708	FORCE
	807	THEOR
402	382	EXCITON
	402	LATTICE
403	293	SYNCHRON
	329	AMBIENT
	357	CONVERT
	403	LEADING
404	89	ENGINEER
	404	MACHINE

404	435	PROGRAM
	484	ACCESS
	511	COMPUT
	632	SYSTEM
	687	DIGIT
	854	DRUM
	938	STOR
405	103	FERROMAG
	166	MAGNETIC
	385	FERRITE
	405	MAGNETI
406	44	COLLISIO
	100	EXTRAORD
	123	HYDROOYN
	231	PROPAGAT
	406	MAGNETO
	480	VELOCIT
	587	PLASMA
	725	IONIC
	726	IONIZ
	876	GYRO
407	67	DEMODULA
	407	MODULAT
408	408	MOLECUL
	657	AMMON
409	409	MULTIPL
410	203	OSCILLAT
	255	RESONATO
	410	NATURAL
411	224	POSITIVE
	411	NEGATIV
412	4	ADMITTAN
	233	QUAORIPO
	294	SYNTHESI
	298	TERMINAL
	311	TRANSFER
	3,0	CANONIC
	412	NETWORK
	498	BRANCH
	555	ITERAT
	556	LADDER
	581	PASSIV
	739	MATRI
413	44	COLLISIO
	87	ELECTRON
	131	INCOHERE
	413	NEUTRAL
	524	DIFFUS
	587	PLASMA
	621	STATIC
	726	IONIZ
	976	GAS
	979	ION
414	414	NEUTRON
	515	COSMIC
	590	PROTON
	681	OECAY
	723	INNER
	834	BELT
	987	RAY
415	234	QUAORUPO
	254	RESONANC
	366	DETONAT
	383	EXPLOSI
	415	NUCLEAR
	553	ISLAND
416	21	ATMOSPHE
	C1	ECCENTRI
	260	SATELLIT
	416	ORBITAL
	424	PERIGEE
	693	EARTH
416	755	ORBIT
	850	DRAG
417	189	MULTISTA
	257	RESPONSE
	256	RESTRICT
	313	TRANSIEN
	417	OVERSHO
	462	STAGGER
	718	IDEAL
	922	RISE
	937	STEP
	942	TIME
418	242	REFLECTO
	418	PARABOL
419	292	SUPERREG
	419	PARAMET
	704	FIGUR
	719	IMAGE
	915	PUMP
	978	IDL
420	420	PARTIAL
	530	ECLIPS
	726	IONIZ
421	3	ADIABATI
	343	CENTRAL
	421	PASSAGE
	698	ERUPT
422	422	PATTERN
423	423	PENTODE
	461	STABILI
	491	AMPLIF
	639	TRIODE
	814	VALVE
	827	ANDD
	875	GRID
424	227	PRESSURE
	260	SATELLIT
	416	ORBITAL
	424	PERIGEE
	540	GRAVIT
	544	HEIGHT
	755	ORBIT
	850	DRAG
425	12	ANISOTRO
	70	DIELECTR
	425	PERMEAB
426	426	PERSIST
427	427	PERTURB
	570	MOTION
428	428	PICTURE
429	429	PLASTIC
	625	STRESS
	764	PRINT
	825	ADHE
	836	BOND
	950	VIEW
430	99	EXTINCTI
	430	POLARIZ
	694	ELLIP
431	431	PREDICT
432	262	SECONDAR
	432	PRIMARY
	532	ENERGY
433	29	BANDWIDT
	433	PRODUCT
434	87	ELECTRON
434	434	PROFILE
	481	VIRTUAL
	519	DENSIT
	544	HEIGHT
435	89	ENGINEER
	404	MACHINE
	435	PROGRAM
	511	COMPUT
436	252	RESISTOR
	436	PROTECT
	792	SILIC
437	437	QUALITY
	914	PULL
438	438	RADIANT
	567	METEOR
	612	SHOWER
	810	TRAIL
439	135	INDUCTOR
	237	REACTANC
	251	RESISTIV
	299	TERMINAT
	439	REACTIV
	498	BRANCH
	990	ROW
440	345	CIRCUIT
	440	REACTOR
	450	SATURAB
	451	SATURAT
	579	OUTPUT
	589	PRESET
	845	CORE
	951	VOLT
441	24	AUTOMATI
	136	INFORMAT
	364	DECIMAL
	441	READING
	589	PRESET
	665	BINAR
	854	DRUM
	919	READ
	940	TAPE
	959	WRIT
	968	COD
442	442	RECOGNI
443	443	RECOVER
	942	TIME
444	99	EXTINCTI
	259	ROUGHNES
	384	FARADAY
	391	INCIDEN
	444	REFLECT
	481	VIRTUAL
	576	OBLIQU
445	236	RADIOSON
	445	REFRACT
	505	CLIMAT
	721	INOEX
446	307	THYRATRO
	446	REGULAT
	468	SUPPLIE
	579	OUTPUT
	763	POWER
	852	DRIV
	951	VOLT
	995	TUN
	998	DC
447	447	RELEASE
	605	ROCKET
	672	CLOUD
448	448	REVERSA

448	692	DRIFT
449	449	REVERSE
	599	RECTIF
	781	SELEN
	792	SILIC
450	307	THYRATRO
	440	REACTOR
	450	SATURAB
	451	SATURAT
	552	INVERT
	845	CORE
451	166	MAGNETIC
	440	REACTOR
	450	SATURAB
	451	SATURAT
452	63	CORRUGAT
	99	EXTINCTI
	131	INCOHERE
	452	SCATTER
	494	ASPECT
	776	RIGID
453	453	SECTION
	677	CROSS
454	15	APPROXIM
	257	RESPONSE
	454	SEGMENT
	559	LINEAR
	832	BASE
455	356	CONTROL
	455	SENSING
456	218	PHOTOSEN
	345	CIRCUIT
	456	SHAPING
	766	PULSE
457	457	SIMILAR
458	458	SIMULAT
459	80	DISTRIBU
	459	SPATIAL
460	17	ARTIFICI
	260	SATELLIT
	373	DOPPLER
	460	SPUTNIK
	537	FADING
	755	ORBIT
461	48	COMPENSA
	101	FEEDBACK
	301	THERMIST
	423	PENTODE
	461	STABILI
	468	SUPPLIE
	499	BRIDGE
	543	HEATER
	875	GRID
	935	STAB
462	491	AMPLIF
	572	NARROW
	633	TABULA
	691	DOUBL
	831	BAND
	905	PASS
	995	TUN
463	113	GEOGRAPH
	120	HEMISPHE
	156	IONOSPHE
	362	DAYTIME
	463	STATION
	752	NORTH
	797	SOUTH
	822	WORLD
463	960	YEAR
464	464	STRAIGH
	888	LINE
465	465	STRATIF
466	196	NONSEASO
	466	SUNRISE
	628	SUNSET
	728	LAYER
	751	NIGHT
467	2	ACTIVITY
	285	STATISTI
	343	CENTRAL
	467	SUNSPOT
	575	NUMBER
	678	CYCLE
	794	SOLAR
	960	YEAR
468	307	THYRATRO
	446	REGULAT
	461	STABILI
	468	SUPPLIE
	579	OUTPUT
	763	POWER
	951	VOLT
	995	TUN
	998	DC
469	63	CORRUGAT
	129	IMPERFEC
	259	ROUGHNES
	336	BARRIER
	469	SURFACE
	513	CONVEX
	923	ROCK
470	470	SYMMETR
471	257	RESPONSE
	378	EQUALIZ
	471	TELEVIS
	508	COLOUR
	831	BAND
472	472	THERMAL
	878	HEAT
	981	LAW
473	163	LIGHTNIN
	473	THUNDER
	706	FLASH
474	164	LONGITUD
	170	MAGNETOS
	474	TORSION
	867	FILT
475	475	TRANSIT
	942	TIME
476	476	TRANSMI
477	32	BISTABLE
	33	BLOCKING
	183	MONOSTAB
	190	MULTIVIB
	314	TRANSIST
	345	CIRCUIT
	477	TRIGGER
	766	PULSE
	869	FLIP
478	108	FREQUENC
	478	TUNABLE
	831	BAND
	917	RANG
479	479	UNIFORM
	621	STATIC
480	406	MAGNETO
	480	VELOCIT
	571	MOVING
	692	DRIFT
	805	SURGE
481	14	APPARENT
	156	IONOSPHE
	272	SOUNDING
	434	PROFILE
	444	REFLECT
	481	VIRTUAL
	544	HEIGHT
	645	VALLEY
	863	FALL
482	37	CENTIMET
	482	VISIBLE
483	483	WINDING
	845	CORE
484	136	INFORMAT
	341	CAPACIT
	404	MACHINE
	484	ACCESS
	595	RANDOM
	735	LOGIC
	742	MEMOR
	854	DRUM
	938	STOR
	966	BIT
485	485	ACCURA
486	486	ACTION
	737	MASER
	795	SOLID
487	128	IMPEDANC
	487	ACTIVE
488	108	FREQUENC
	488	ADJUST
	497	BALANC
	543	HEATER
	814	VALVE
489	184	MORPHOLO
	489	ADVANC
	529	DYNAMO
490	75	DIRECTIV
	145	INTERFER
	253	RESOLUTI
	295	TELESCOP
	333	ANTENNA
	490	AERIAL
	523	DIAMET
	596	RECEIV
	600	RESOLV
	660	ARRAY
	661	ASTRO
491	29	BANDWIDT
	48	COMPENSA
	292	SUPERREG
	377	EMITTER
	423	PENTODE
	462	STAGGER
	491	AMPLIF
	497	BALANC
	704	FIGUR
	838	CHOP
	874	GAIN
	915	PUMP
	978	IDL
492	196	NONSEASO
	492	ANOMAL
	930	SKIN
493	156	IONOSPHE
	493	ARCTIC

494	157	IRREGULA
	266	SENSITIV
	452	SCATTER
	494	ASPECT
	655	ALIGN
	858	ECHO
495	191	NITROGEN
	240	RECOMBIN
	495	ATOMIC
	580	OXYGEN
	829	ATOM
496	269	SIMULTAN
	496	AURORA
	553	ISLAND
	965	BAY
	996	ZON
497	28	BALANCED
	345	CIRCUIT
	488	ADJUST
	491	AMPLIF
	497	BALANC
	543	HEATER
	552	INVERT
	579	OUTPUT
	814	VALVE
	951	VOLT
	998	DC
498	128	IMPEDANC
	138	INSERTIO
	233	QUADRIPO
	299	TERMINAT
	412	NETWORK
	439	REACTIV
	498	BRANCH
	556	LADDER
	719	IMAGE
	867	FILT
	905	PASS
499	101	FEEDBACK
	263	SELECTIV
	461	STABILI
	499	BRIDGE
	855	DUAL
500	80	DISTRIBU
	200	OBSERVAT
	235	RADIATIO
	500	BRIGHT
	514	CORONA
	573	NEBULA
	771	RADIO
	828	AREA
	964	ARC
	993	SUN
501	252	RESISTOR
	501	CARBON
	535	EXTREM
	866	FILM
502	128	IMPEDANC
	233	QUADRIPO
	502	CASCAD
	905	PASS
503	503	CATHOD
	827	ANOD
	842	COLD
504	223	POSITION
	504	CENTRE
505	445	REFRACT
	505	CLIMAT
506	279	SPONTANE
	506	COHERE
507	507	CLOSED
	632	SYSTEM
	782	SERVO
	892	LOOP
508	471	TELEVIS
	508	COLOUR
509	509	COLUMN
510	510	COMBIN
	559	LINEAR
511	511	COMPOS
512	10	ANALOGUE
	89	ENGINEER
	136	INFORMAT
	404	MACHINE
	435	PROGRAM
	512	COMPUT
	687	DIGIT
	854	DRUM
	919	READ
	940	TAPE
513	20	ASYMPTOT
	72	DIFFRACT
	350	CONDUCT
	395	INFINIT
	469	SURFACE
	513	CONVEX
	516	CYLIND
	759	PLANE
	967	BOD
514	235	RADIATIO
	500	BRIGHT
	514	CORONA
	698	ERUPT
	794	SOLAR
	993	SUN
515	143	INTENSIT
	149	INTERPLA
	300	TERRESTR
	335	BALLOON
	414	NEUTRON
	515	COSMIC
	578	ORIGIN
	590	PROTON
	705	FLARE
	834	BELT
	871	FLUX
	987	RAY
516	1	ACOUSTIC
	277	SPHERICA
	350	CONDUCT
	513	CONVEX
	516	CYLIND
	617	SPHERE
	662	AXIAL
	673	CONIC
	786	SHEET
	830	AXES
517	517	DEFORM
518	518	DEGREE
519	52	CONCENTR
	196	NONSEASO
	384	FARADAY
	434	PROFILE
	519	DENSIT
	524	DIFFUS
	645	VALLEY
	780	SCALE
	813	UPPER
	863	FALL
520	23	ATTENUAT
	175	MICROMIN
520	263	SELECTIV
	281	SPURIOUS
	308	TOLERANC
	378	EQUALIZ
	520	DESIGN
	663	BASIC
	691	DOUBL
	738	MATCH
521	521	DETECT
522	522	DEVIAT
523	145	INTERFER
	295	TELESCOP
	490	AERIAL
	523	DIAMET
	568	MIRROR
	661	ASTRO
	744	METRE
524	22	ATTACHME
	87	ELECTRON
	153	IONIZATI
	413	NEUTRAL
	519	DENSIT
	524	DIFFUS
	726	IONIZ
	979	ION
525	525	DIRECT
526	526	DISTAN
527	527	DIVERG
528	528	DOMAIN
529	114	GEOMAGNE
	117	GRADIENT
	156	IONOSPHE
	265	SEMIDIUR
	489	ADVANC
	529	DYNAMO
	707	FLUID
	768	QUIET
	807	THEOR
	894	MAIN
	939	TAIL
	994	TID
530	332	ANNULAR
	420	PARTIAL
	530	ECLIPS
	794	SOLAR
531	262	SECONDAR
	279	SPONTANE
	338	BOMBARD
	531	EMISSI
	623	STIMUL
	634	TARGET
	860	EMIT
532	93	EXCHANGE
	144	INTERACT
	182	MOMENTUM
	262	SECONDAR
	351	CONSERV
	401	KINETIC
	432	PRIMARY
	532	ENERGY
533	533	EXAMIN
534	534	EXTEND
535	204	OSCILLOG
	501	CARBON
	535	EXTREM
	604	RIPPLE
	935	STAB
	982	LOW

536	536	FACTOR
537	157	IRREGULA
	384	FARADAY
	460	SPUTNIK
	537	FADING
	613	SIGNAL
	956	WIND
538	20	ASYMPTOT
	172	MATHEMAT
	271	SOLUTION
	284	STATIONA
	339	BOUNDAR
	538	FINITE
	643	UNIQUE
	647	VECTOR
	796	SOLVE
539	539	FLIGHT
	557	LAUNCH
	563	MEASUR
	605	ROCKET
540	424	PERIGEE
	540	GRAVIT
	693	EARTH
541	541	GROUND
542	246	RELATIVI
	284	STATIONA
	542	GROWTH
543	28	BALANCED
	461	STABILI
	488	ADJUST
	497	BALANC
	543	HEATER
	814	VALVE
	951	VOLT
544	31	BIFURCAT
	196	NONSEASO
	280	SPORADIC
	325	VERTICAL
	424	PERIGEE
	434	PROFILE
	481	VIRTUAL
	544	HEIGHT
	628	SUNSET
	751	NIGHT
	813	UPPER
545	191	NITROGEN
	227	PRESSURE
	387	GASEOUS
	545	HELIUM
	587	PLASMA
	765	PROBE
546	59	CONVENTI
	160	JUNCTION
	546	HYBRID
547	547	IDENTI
548	548	IMPACT
549	549	IMPULS
550	550	INCLIN
	755	ORBIT
551	246	RELATIVI
	551	INJECT
552	450	SATURAB
	497	BALANC
	552	INVERT
	579	OUTPUT
	852	DRIV
	908	PHAS
553	383	EXPLOSI
	415	NUCLEAR
	496	AURORA
	553	ISLAND
554	554	ISOLAT
555	412	NETWORK
	555	ITERAT
556	128	IMPEDANC
	138	INSERTIO
	299	TERMINAT
	412	NETWORK
	498	BRANCH
	556	LADDER
	719	IMAGE
	738	MATCH
	867	FILT
	905	PASS
557	260	SATELLIT
	539	FLIGHT
	557	LAUNCH
	605	ROCKET
558	558	LENGTH
559	313	TRANSIEN
	454	SEGMENT
	510	COMBIN
	559	LINEAR
	739	MATRI
560	560	LIQUID
561	561	MAGNET
562	562	MATTER
563	168	MAGNETOM
	236	RADIOSON
	259	ROUGHNES
	539	FLIGHT
	563	MEASUR
	605	ROCKET
	765	PROBE
564	6	AFTERNOO
	564	MEDIAN
	678	CYCLE
	679	DAILY
	749	MONTH
	960	YEAR
565	12	ANISOTRO
	70	DIELECTR
	86	ELECTROM
	565	MEDIUM
	741	MEDIA
	954	WAVE
566	76	DISCHARG
	566	MERCUR
567	319	TURBULEN
	438	RADIANT
	567	METEOR
	612	SHOWER
	810	TRAIL
	858	ECHO
568	295	TELESCOP
	523	DIAMET
	568	MIRROR
	734	LOCAT
569	144	INTERACT
	166	MAGNETIC
	167	MAGNETIS
	569	MOMENT
	703	FIELD
	722	INDUC
	934	SPIN
570	427	PERTURB
	570	MOTION
571	38	CERENKOV
	208	PARTICLE
	246	RELATIVI
	480	VELOCIT
	571	MOVING
572	29	BANDWIDT
	108	FREQUENC
	263	SELECTIV
	462	STAGGER
	572	NARROW
	676	COUPL
	831	BAND
573	78	DISCRETE
	111	GALACTIC
	200	OBSERVAT
	235	RADIATIO
	500	BRIGHT
	573	NEBULA
	615	SOURCE
	661	ASTRO
	709	GALAX
	771	RADIO
	964	ARC
574	574	NORMAL
575	467	SUNSPOT
	575	NUMBER
576	99	EXTINCTI
	391	INCIDEN
	444	REFLECT
	576	OBLIQU
577	99	EXTINCTI
	577	ORIENT
578	515	COSMIC
	578	ORIGIN
579	83	EFFICIEN
	307	THYRATRO
	440	REACTOR
	446	REGULAT
	468	SUPPLIE
	497	BALANC
	552	INVERT
	579	OUTPUT
	724	INPUT
	852	DRIV
	962	ADD
580	21	ATMOSPHE
	42	COEFFICI
	52	CONCENTR
	106	FORMATIO
	191	NITROGEN
	238	REACTION
	240	RECOMBIN
	495	ATOMIC
	580	OXYGEN
	710	GASES
	863	FALL
581	4	ADMITTAN
	195	NONRECIP
	233	QUADRIPO
	239	RECIPROC
	298	TERMINAL
	340	CANONIC
	412	NETWORK
	581	PASSIV
	739	MATRI
582	178	MIDNIGHT
	285	STATISTI
	343	CENTRAL
	582	PERIOD

583	583	PHOTON
584	584	PHYSIC
585	585	PLANAR
586	400	JUPITER
	586	PLANET
587	66	CYCLOTRO
	123	HYDRODYN
	164	LONGITUD
	283	STANDING
	381	EXCITED
	387	GASEOUS
	406	MAGNETO
	413	NEUTRAL
	545	HELIUM
	587	PLASMA
	707	FLUID
	765	PROBE
588	356	CONTROL
	588	PRECIS
	890	LOCK
589	356	CONTROL
	440	REACTOR
	441	READING
	589	PRESET
	598	RECTAN
	613	SIGNAL
	800	SPEED
	864	FAST
	890	LOCK
	966	BIT
	977	GAT
590	208	PARTICLE
	414	NEUTRON
	515	COSMIC
	590	PROTON
	723	INNER
	757	OUTER
	834	BELT
	944	TRAP
	987	RAY
591	114	GEOMAGNE
	125	HYDROMAG
	176	MICROPUL
	591	PULSAT
	712	GIANT
	803	STORM
	965	BAY
592	592	RADIAL
	770	RADII
593	593	RADIAT
594	594	RADIUS
	967	BOD
595	286	STOCHAST
	484	ACCESS
	595	RANDOM
596	217	PHOTOGRA
	295	TELESCOP
	333	ANTENNA
	490	AERIAL
	596	RECEIV
597	18	ASSOCIAT
	150	INTERVAL
	155	IONOSOND
	217	PHOTOGRA
	280	SPORADIC
	597	RECORD
	645	VALLEY
	712	GIANT
	877	HEAD
597	940	TAPE
598	589	PRESET
	598	RECTAN
599	116	GERMANIU
	449	REVERSE
	599	RECTIF
	763	POWER
	781	SELEN
	792	SILIC
	844	COOL
600	253	RESOLUTI
	490	AERIAL
	600	RESOLV
	763	POWER
601	47	COMPATIB
	108	FREQUENC
	203	OSCILLAT
	601	RESONA
602	602	RETARD
603	163	LIGHTNIN
	361	CURRENT
	603	RETURN
	626	STROKE
604	535	EXTREM
	604	RIPPLE
605	140	INSTRUME
	447	RELEASE
	539	FLIGHT
	557	LAUNCH
	563	MEASUR
	605	ROCKET
606	20	ASYMPTOT
	40	CIRCULAR
	72	DIFFRACT
	271	SOLUTION
	278	SPHEROID
	395	INFINIT
	606	SCALAR
	643	UNIQUE
	647	VECTOR
	776	RIGID
607	72	DIFFRACT
	86	ELECTROM
	271	SOLUTION
	350	CONDUCT
	395	INFINIT
	607	SCREEN
	759	PLANE
	931	SLIT
	954	WAVE
608	13	APPARATU
	608	SEARCH
609	156	IONOSPHE
	196	NONSEASO
	297	TEMPORAL
	324	VARIATIO
	362	DAYTIME
	372	DIURNAL
	609	SEASON
	627	SUMMER
	652	WINTER
	678	CYCLE
	679	DAILY
	728	LAYER
	749	MONTH
	960	YEAR
610	183	MONOSTAB
	281	SPURIOUS
	326	WAVEFORM
	610	SELECT
	636	TIMING
610	683	DELAY
	766	PULSE
	793	SINGL
	864	FAST
611	611	SERIES
612	438	RADIANT
	567	METEOR
	612	SHOWER
	649	VISUAL
	810	TRAIL
613	537	FADING
	589	PRESET
	613	SIGNAL
	978	IDL
	984	MIX
614	614	SILVER
	836	BOND
615	78	DISCRETE
	111	GALACTIC
	145	INTERFER
	573	NEBULA
	615	SOURCE
	629	SURVEY
616	286	STOCHAST
	616	SPECTR
617	86	ELECTROM
	350	CONDUCT
	516	CYLIND
	617	SPHERE
	807	THEOR
	830	AXES
618	618	SPLITT
619	155	IONOSOND
	156	IONOSPHE
	157	IRREGULA
	619	SPREAD
	655	ALIGN
	858	ECHO
620	620	SQUARE
621	44	COLLISIO
	66	CYCLOTRO
	121	HOMOGENE
	131	INCOHERE
	316	TRANSPOR
	317	TRANSVER
	394	INERTIA
	413	NEUTRAL
	479	UNIFORM
	621	STATIC
	976	GAS
622	172	MATHEMAT
	386	FOURIER
	622	STEADY
623	531	EMISSI
	623	STIMUL
	737	MASER
624	624	STREAM
625	429	PLASTIC
	625	STRESS
	825	ADHE
	836	BOND
626	163	LIGHTNIN
	290	SUCCESSI
	603	RETURN
	626	STROKE
	706	FLASH
627	156	IONOSPHE

627	324	VARIATIO
	362	DAYTIME
	609	SEASON
	627	SUMMER
	652	WINTER
	679	DAILY
	728	LAYER
	797	SOUTH
	969	DAY
628	466	SUNRISE
	544	HEIGHT
	628	SUNSET
	848	DATA
629	615	SOURCE
	629	SURVEY
630	30	BIDIRECT
	32	BISTABLE
	630	SWITCH
	665	8INAR
	735	LOGIC
	774	RELAY
	869	FLIP
	977	GAT
631	172	MATHEMAT
	631	SYMBOL
	632	SYSTEM
632	24	AUTOMATI
	136	INFORMAT
	404	MACHINE
	507	CLOSED
	631	SYMBOL
	632	SYSTEM
	782	SERVO
	830	AXES
	966	8IT
633	198	NORMALIZ
	462	STAGGER
	633	TABULA
	719	IMAGE
634	87	ELECTRON
	262	SECONDAR
	531	EMISSI
	634	TARGET
635	12	ANISOTRO
	70	DIELECTR
	121	HOMOGENE
	346	CLASSIC
	399	ISOTROP
	635	TENSOR
	703	FIELD
	741	MEDIA
636	33	BLOCKING
	610	SELECT
	636	TIMING
	864	FAST
637	637	TOROID
	845	CORE
638	638	TRIANG
639	5	ADVANTAG
	423	PENTODE
	639	TRIODE
	642	TUNNEL
	814	VALVE
	827	ANOD
	875	GRID
	935	STAB
640	640	TRIPLE
	691	DOUBL
	641	TROUGH
	728	LAYER
642	357	CONVERT
	639	TRIODE
	642	TUNNEL
	689	DIODE
	874	GAIN
	935	STAB
	978	IDL
643	20	ASYMPTOT
	271	SOLUTION
	339	8OUNDAR
	386	FOURIER
	538	FINITE
	606	SCALAR
	643	UNIQUE
644	87	ELECTRON
	94	EXCITATI
	130	IMPURITY
	350	CONDUCT
	644	VALENC
	829	ATOM
	831	8AND
645	100	EXTRAORD
	154	IONOGRAM
	481	VIRTUAL
	519	DENSIT
	597	RECORD
	645	VALLEY
	846	CURV
	907	PEAK
646	646	VAPOUR
	817	WATER
647	40	CIRCULAR
	172	MATHEMAT
	212	PERPENDI
	538	FINITE
	606	SCALAR
	647	VECTOR
	703	FIELD
648	360	CRYSTAL
	382	EXCITON
	648	VI8RAT
	785	SHEAR
	898	MODE
649	200	OBSERVAT
	217	PHOTOGRA
	612	SHOWER
	649	VISUAL
	858	ECHO
650	650	VOLUME
651	95	EXOSPHER
	651	WHISTL
	856	DUCT
652	156	IONOSPHE
	324	VARIATIO
	362	DAYTIME
	609	SEASON
	627	SUMMER
	652	WINTER
	678	CYCLE
	679	DAILY
	728	LAYER
	969	DAY
653	653	A8SOR
654	654	ADAPT
655	157	IRREGULA
	494	ASPECT
	619	SPREAD
	655	ALIGN
	856	DUCT
656	160	JUNCTION
	656	ALLOY
657	408	MOLECUL
	657	AMMON
	737	MASER
	888	LINE
658	75	DIRECTIV
	658	ANALY
	883	JUMP
659	659	ANGLE
	818	WEDGE
660	490	AERIAL
	660	ARRAY
661	145	INTERFER
	200	OBSERVAT
	295	TELESCOP
	490	AERIAL
	523	DIAMET
	573	NE8ULA
	661	ASTRO
	771	RADIO
	993	SUN
662	516	CYLIND
	662	AXIAL
663	201	OPERATIO
	314	TRANSIST
	520	DESIGN
	663	BASIC
	665	8INAR
	687	DIGIT
	735	LOGIC
664	664	8ASIS
665	258	REVERSI8
	364	DECIMAL
	441	READING
	630	SWITCH
	663	BASIC
	665	BINAR
	687	DIGIT
	735	LOGIC
	962	ADD
	977	GAT
666	666	8LOCK
667	205	OUTBURST
	667	8URST
	698	ERUPT
	705	FLARE
	868	FINE
668	255	RESONATO
	381	EXCITED
	668	CAVIT
669	208	PARTICLE
	246	RELATIVI
	309	TRAJECTO
	669	CHARG
	798	SPACE
	896	MASS
670	670	CIRCL
671	314	TRANSIST
	370	DISTORT
	377	EMITTER
	671	CLASS
	914	PULL
672	21	ATMOSPHE
	163	LIGHTNIN
	447	RELEASE
	672	CLOUD

673	55	CONFIGUR
	339	BOUNDAR
	516	CYLIND
	673	CONIC
	696	EQUAT
	772	RATIO
	840	COAX
674	674	CONNE
675	218	PHOTOSEN
	258	REVERSIB
	675	COUNT
	766	PULSE
	962	ADD
676	572	NARROW
	676	COUPL
	691	DOUBL
677	453	SECTION
	677	CROSS
678	2	ACTIVITY
	6	AFTERNOO
	467	SUNSPOT
	564	MEDIAN
	609	SEASON
	652	WINTER
	678	CYCLE
	749	MONTH
	960	YEAR
	969	DAY
679	113	GEOGRAPH
	156	IONOSPHE
	162	LATITUDE
	196	NONSEASO
	324	VARIATIO
	372	DIURNAL
	564	MEDIAN
	609	SEASON
	627	SUMMER
	652	WINTER
	679	DAILY
	728	LAYER
680	110	FUNDAMEN
	203	OSCILLAT
	680	DAMPI
681	414	NEUTRON
	681	DECAY
682	682	DEFIN
683	610	SELECT
	683	DELAY
684	684	DENSE
685	209	PENETRAT
	685	DEPTH
	903	OXID
	930	SKIN
686	686	DERIV
	746	MOBIL
687	136	INFORMAT
	364	DECIMAL
	404	MACHINE
	511	COMPUT
	663	BASIC
	665	BINAR
	687	DIGIT
	854	DRUM
	940	TAPE
	962	ADD
	966	BIT
	968	COD
688	688	DILUT
689	32	BISTABLE
	292	SUPERREG
	642	TUNNEL
	689	DIODE
690	690	DIPOL
	722	INDUC
691	462	STAGGER
	520	DESIGN
	640	TRIPLE
	676	COUPL
	691	DOUBL
692	448	REVERSA
	480	VELOCIT
	692	DRIFT
693	17	ARTIFICI
	95	EXOSPHER
	149	INTERPLA
	260	SATELLIT
	300	TERRESTR
	416	ORBITAL
	540	GRAVIT
	693	EARTH
	755	ORBIT
	757	OUTER
	770	RADII
	834	BELT
	894	MAIN
694	40	CIRCULAR
	99	EXTINCTI
	430	POLARIZ
	694	ELLIP
	759	PLANE
695	239	RECIPROC
	695	EQUAL
696	141	INTEGRAL
	172	MATHEMAT
	271	SOLUTION
	284	STATIONA
	673	CONIC
	696	EQUAT
	700	EXPAN
	796	SOLVE
697	697	ERROR
698	205	OUTBURST
	421	PASSAGE
	514	CORONA
	667	BURST
	698	ERUPT
	705	FLARE
699	20	ASYMPTOT
	129	IMPERFEC
	271	SOLUTION
	350	CONDUCT
	395	INFINIT
	699	EXACT
700	15	APPROXIM
	20	ASYMPTOT
	72	DIFFRACT
	141	INTEGRAL
	271	SOLUTION
	278	SPHEROID
	696	EQUAT
	700	EXPAN
	818	WEDGE
701	701	FAULT
702	254	RESONANC
	702	FERRI
	934	SPIN
703	12	ANISOTRO
	63	CORRUGAT
703	121	HOMOGENE
	169	MAGNETOR
	287	STRENGTH
	316	TRANSPOR
	317	TRANSVER
	346	CLASSIC
	569	MOMENT
	635	TENSOR
	647	VECTOR
	703	FIELD
	707	FLUID
	708	FORCE
	711	GAUSS
704	29	BANDWIDT
	419	PARAMET
	491	AMPLIF
	704	FIGUR
	874	GAIN
	900	NOIS
	915	PUMP
	978	IDL
	982	LOW
705	205	OUTBURST
	358	CORPUSC
	515	COSMIC
	667	BURST
	698	ERUPT
	705	FLARE
	736	MAJOR
	794	SOLAR
	803	STORM
	987	RAY
706	163	LIGHTNIN
	473	THUNDER
	626	STROKE
	706	FLASH
707	121	HOMOGENE
	125	HYDROMAG
	166	MAGNETIC
	350	CONDUCT
	529	DYNAMO
	587	PLASMA
	703	FIELD
	707	FLUID
708	166	MAGNETIC
	208	PARTICLE
	246	RELATIVI
	401	KINETIC
	703	FIELD
	708	FORCE
	717	HELIC
	870	FLOW
709	111	GALACTIC
	235	RADIATIO
	573	NEBULA
	709	GALAX
	771	RADIO
710	87	ELECTRON
	191	NITROGEN
	227	PRESSURE
	580	OXYGEN
	710	GASES
	726	IONIZ
	976	GAS
	979	ION
711	166	MAGNETIC
	703	FIELD
	711	GAUSS
712	114	GEOMAGNE
	371	DISTURB
	591	PULSAT
	597	RECORD
	712	GIANT
	976	GAS

713	713	GLASS
	879	HIGH
714	714	GRAPH
715	715	GROUP
716	716	HEAVY
717	708	FORCE
	717	HELIC
	888	LINE
718	417	OVERSHO
	718	IDEAL
	719	IMAGE
719	128	IMPEDANC
	138	INSERTIO
	233	QUADRIPO
	419	PARAMET
	498	BRANCH
	556	LADDER
	633	TABULA
	718	IDEAL
	719	IMAGE
	738	MATCH
	867	FILT
	905	PASS
720	720	IMPUR
	743	METAL
	916	PURE
721	236	RADIOSON
	445	REFRACT
	721	INDEX
722	361	CURRENT
	569	MOMENT
	690	DIPOL
	722	INDUC
723	414	NEUTRON
	590	PROTON
	723	INNER
	757	OUTER
724	293	SYNCHRON
	313	TRANSIEN
	357	CONVERT
	579	OUTPUT
	724	INPUT
	838	CHOP
	852	DRIV
	865	FEED
	968	COD
725	44	COLLISIO
	123	HYDRODYN
	406	MAGNETO
	725	IONIC
	726	IONIZ
726	123	HYDRODYN
	316	TRANSPOR
	406	MAGNETO
	413	NEUTRAL
	420	PARTIAL
	524	DIFFUS
	710	GASES
	725	IONIC
	726	IONIZ
	976	GAS
727	727	JOINT
728	6	AFTERNOO
	31	BIFURCAT
	196	NONSEASO
	325	VERTICAL
	379	EQUATOR
	466	SUNRISE
	609	SEASON
728	627	SUMMER
	641	TROUGH
	652	WINTER
	679	DAILY
	728	LAYER
	882	HOUR
729	729	LEADS
730	3	ADIABATI
	730	LEVEL
	737	MASER
731	731	LIGHT
732	732	LIMIT
733	114	GEOMAGNE
	178	MIDNIGHT
	371	DISTURB
	372	DIURNAL
	733	LOCAL
734	366	DETONAT
	568	MIRROR
	734	LOCAT
735	175	MICROMIN
	201	OPERATIO
	484	ACCESS
	630	SWITCH
	663	BASIC
	665	BINAR
	735	LOGIC
	966	BIT
736	200	OBSERVAT
	358	CORPUSC
	705	FLARE
	736	MAJOR
	857	EAST
737	3	ADIABATI
	222	POPULATI
	234	QUADRUPO
	486	ACTION
	623	STIMUL
	657	AMMON
	730	LEVEL
	737	MASER
	795	SOLID
	915	PUMP
738	128	IMPEDANC
	299	TERMINAT
	520	DESIGN
	556	LADDER
	719	IMAGE
	738	MATCH
	867	FILT
	905	PASS
739	4	ADMITTAN
	188	MULTIPOL
	248	REPRESEN
	340	CANONIC
	412	NETWORK
	559	LINEAR
	581	PASSIV
	739	MATRI
740	64	CRITICAL
	178	MIDNIGHT
	740	MAXIM
741	12	ANISOTRO
	70	DIELECTR
	86	ELECTROM
	231	PROPAGAT
	565	MEDIUM
	635	TENSOR
	741	MEDIA
	954	WAVE
742	43	COINCIDE
	65	CRYOTRON
	136	INFORMAT
	484	ACCESS
	742	MEMOR
	938	STOR
	966	BIT
743	92	EVAPORAT
	216	PHOTOELE
	262	SECONDAR
	397	INSULAT
	720	IMPUR
	743	METAL
	816	WAFER
	825	ADHE
	903	OXID
	930	SKIN
	941	THIN
	948	VACU
744	328	WAVELENG
	523	DIAMET
	744	METRE
745	745	MINIM
746	84	ELECTRIC
	316	TRANSPOR
	686	DERIV
	746	MOBIL
747	747	MODEL
	939	TAIL
748	175	MICROMIN
	748	MODUL
749	564	MEDIAN
	609	SEASON
	678	CYCLE
	749	MONTH
	897	MEAN
750	356	CONTROL
	750	MOTOR
	852	DRIV
751	362	DAYTIME
	466	SUNRISE
	544	HEIGHT
	751	NIGHT
	780	SCALE
	863	FALL
752	120	HEMISPHE
	156	IONOSPHE
	463	STATION
	752	NORTH
	762	POLAR
	797	SOUTH
	965	BAY
753	218	PHOTOSEN
	753	OPTIC
754	29	BANDWIDT
	754	OPTIM
755	17	ARTIFICI
	81	ECCENTRI
	260	SATELLIT
	416	ORBITAL
	424	PERIGEE
	460	SPUTNIK
	550	INCLIN
	693	EARTH
	755	ORBIT
756	756	ORDER
757	208	PARTICLE
	590	PROTON
	693	EARTH

757	723	INNER
	757	OUTER
	834	BELT
758	758	PAPER
759	72	DIFFRACT
	73	DIMENSIO
	212	PERPENDI
	277	SPHERICA
	339	BOUNDAR
	395	INFINIT
	513	CONVEX
	607	SCREEN
	694	ELLIP
	759	PLANE
	818	WEDGE
	931	SLIT
760	760	PLATE
	785	SHEAR
761	353	CONTACT
	761	POINT
	950	VIEW
762	114	GEOMAGNE
	162	LATITUDE
	371	DISTURB
	752	NORTH
	762	POLAR
	803	STORM
	996	ZON
763	83	EFFICIEN
	446	REGULAT
	468	SUPPLIE
	599	RECTIF
	600	RESOLV
	763	POWER
	792	SILIC
	990	ROW
	998	DC
764	429	PLASTIC
	764	PRINT
	825	ADHE
	836	BOND
765	76	DISCHARG
	87	ELECTRON
	124	HYDROGEN
	227	PRESSURE
	387	GASEOUS
	545	HELIUM
	563	MEASUR
	587	PLASMA
	765	PROBE
	976	GAS
	979	ION
766	180	MILLIMIC
	326	WAVEFORM
	456	SHAPING
	477	TRIGGER
	610	SELECT
	675	COUNT
	766	PULSE
	811	TRAIN
	922	RISE
767	767	QUANT
	807	THEOR
768	162	LATITUDE
	200	OBSERVAT
	372	DIURNAL
	379	EQUATOR
	529	DYNAMO
	768	QUIET
	794	SOLAR
	969	DAY
769	769	RADAR
769	858	ECHO
770	592	RADIAL
	693	EARTH
	770	RADII
	921	RING
771	111	GALACTIC
	145	INTERFER
	295	TELESCOP
	331	ANGULAR
	400	JUPITER
	500	BRIGHT
	573	NEBULA
	661	ASTRO
	709	GALAX
	771	RADIO
	936	STAR
772	673	CONIC
	772	RATIO
773	773	REDUC
774	5	ADVANTAG
	630	SWITCH
	774	RELAY
	855	DUAL
775	20	ASYMPTOT
	81	ECCENTRI
	775	REVOL
776	452	SCATTER
	606	SCALAR
	776	RIGID
777	384	FARADAY
	777	ROTAT
778	778	ROUND
779	779	SAMPL
780	159	ISOTHERM
	319	TURBULEN
	325	VERTICAL
	519	DENSIT
	751	NIGHT
	780	SCALE
781	39	CHARACTE
	449	REVERSE
	599	RECTIF
	781	SELEN
	951	VOLT
782	101	FEEDBACK
	257	RESPONSE
	507	CLOSED
	632	SYSTEM
	782	SERVO
	892	LOOP
783	783	SHAPE
784	784	SHARP
785	305	THICKNES
	360	CRYSTAL
	648	VIBRAT
	760	PLATE
	785	SHEAR
	898	MODE
786	63	CORRUGAT
	212	PERPENDI
	350	CONDUCT
	395	INFINIT
	516	CYLIND
	786	SHEET
787	787	SHELL
788	373	DOPPLER
	788	SHIFT
	908	PHAS
789	789	SHOCK
790	790	SHORT
791	341	CAPACIT
	377	EMITTER
	791	SHUNT
	889	LOAD
792	116	GERMANIU
	160	JUNCTION
	436	PROTECT
	449	REVERSE
	599	RECTIF
	763	POWER
	792	SILIC
793	610	SELECT
	793	SINGL
794	205	OUTBURST
	467	SUNSPOT
	514	CORONA
	530	ECLIPS
	705	FLARE
	768	QUIET
	794	SOLAR
	960	YEAR
	993	SUN
795	486	ACTION
	737	MASER
	795	SOLID
	934	SPIN
796	8	ALTERNAT
	141	INTEGRAL
	271	SOLUTION
	346	CLASSIC
	538	FINITE
	696	EQUAT
	796	SOLVE
	967	BOD
797	113	GEOGRAPH
	114	GEOMAGNE
	120	HEMISPHE
	156	IONOSPHE
	200	OBSERVAT
	372	DIURNAL
	463	STATION
	627	SUMMER
	752	NORTH
	797	SOUTH
798	149	INTERPLA
	669	CHARG
	798	SPACE
	815	VEHIC
799	76	DISCHARG
	799	SPARK
800	136	INFORMAT
	589	PRESET
	800	SPEED
	879	HIGH
	919	READ
	938	STOR
	940	TAPE
	962	ADD
801	152	INVERSIO
	312	TRANSFOR
	801	SPLIT
802	802	START
803	114	GEOMAGNE
	184	MORPHOLO

803	205	OUTBURST
	285	STATISTI
	300	TERRESTR
	591	PULSAR
	705	FLARE
	762	POLAR
	803	STORM
	894	MAIN
	965	BAY
804	804	STRIP
	931	SLIT
805	480	VELOCIT
	805	SURGE
806	108	FREQUENC
	806	SWEEP
807	88	ELECTROS
	103	FERROMAG
	286	STOCHAST
	346	CLASSIC
	382	EXCITON
	401	KINETIC
	529	DYNAMO
	617	SPHERE
	767	QUANT
	807	THEOR
808	100	EXTRAORD
	808	TRACE
809	809	TRACK
810	319	TURBULEN
	438	RADIANT
	567	METEOR
	612	SHOWER
	810	TRAIL
	858	ECHO
811	150	INTERVAL
	766	PULSE
	811	TRAIN
812	812	TWIST
	888	LINE
	898	MODE
813	21	ATMOSPHE
	519	DENSIT
	544	HEIGHT
	813	UPPER
814	423	PENTODE
	488	ADJUST
	497	BALANC
	543	HEATER
	639	TRIODE
	814	VALVE
	827	ANOD
	875	GRID
	995	TUN
	998	DC
815	260	SATELLIT
	798	SPACE
	815	VEHIC
816	39	CHARACTE
	50	COMPONEN
	743	METAL
	816	WAFER
817	646	VAPOUR
	817	WATER
818	1	ACOUSTIC
	72	DIFFRACT
	86	ELECTROM
	141	INTEGRAL
	659	ANGLE
	700	EXPAN
818	759	PLANE
	818	WEDGE
	859	EDGE
819	819	WEIGH
820	820	WHITE
	900	NOIS
821	821	WIDTH
822	113	GEOGRAPH
	162	LATITUDE
	463	STATION
	822	WORLD
823	65	CRYOTRON
	134	INDUCTAN
	252	RESISTOR
	823	WOUND
	841	COIL
	957	WIRE
824	824	YIELD
825	50	COMPONEN
	429	PLASTIC
	625	STRESS
	743	METAL
	764	PRINT
	825	ADHE
	836	BOND
	950	VIEW
826	106	FORMATIO
	335	BALLOON
	826	ALTI
	99[illegible]	X
827	423	PENTODE
	503	CATHOD
	639	TRIODE
	814	VALVE
	827	ANOD
	865	FEED
	875	GRID
828	149	INTERPLA
	500	BRIGHT
	828	AREA
	857	EAST
	993	SUN
829	58	CONSTITU
	87	ELECTRON
	495	ATOMIC
	644	VALENC
	829	ATOM
	909	PHOT
830	12	ANISOTRO
	85	ELECTROD
	88	ELECTROS
	346	CLASSIC
	516	CYLIND
	617	SPHERE
	632	SYSTEM
	830	AXES
831	23	ATTENUAT
	29	BANDWIDT
	138	INSERTIO
	462	STAGGER
	471	TELEVIS
	478	TUNABLE
	572	NARROW
	644	VALENC
	831	BAND
	867	FILT
	905	PASS
832	158	IRREVERS
	454	SEGMENT
	832	BASE
833	289	SUBMILLI
	833	BEAM
834	208	PARTICLE
	235	RADIATIO
	414	NEUTRON
	515	COSMIC
	590	PROTON
	693	EARTH
	757	OUTER
	834	BELT
	944	TRAP
	987	RAY
835	183	MONOSTAB
	314	TRANSIST
	377	EMITTER
	835	BIAS
	875	GRID
836	429	PLASTIC
	614	SILVER
	625	STRESS
	764	PRINT
	825	ADHE
	836	BOND
	950	VIEW
837	837	CELL
838	314	TRANSIST
	491	AMPLIF
	724	INPUT
	838	CHOP
	874	GAIN
	998	DC
839	839	COAT
	903	OXID
840	673	CONIC
	840	COAX
	888	LINE
841	134	INDUCTAN
	823	WOUND
	841	COIL
842	503	CATHOD
	842	COLD
843	843	CONE
844	329	AMBIENT
	599	RECTIF
	844	COOL
845	385	FERRITE
	440	REACTOR
	450	SATURAB
	483	WINDING
	637	TOROID
	845	CORE
846	645	VALLEY
	846	CURV
	880	HISS
847	847	DARK
	939	TAIL
848	236	RADIOSON
	628	SUNSET
	848	DATA
849	244	REGENERA
	388	GENERAT
	849	DIVI
	890	LOCK
850	21	ATMOSPHE
	260	SATELLIT
	416	ORBITAL

850	424	PERIGEE
	850	DRAG
	963	AIR
851	851	DRAW
852	101	FEEDBACK
	345	CIRCUIT
	446	REGULAT
	552	INVERT
	579	OUTPUT
	724	INPUT
	750	MOTOR
	852	DRIV
	865	FEED
	951	VOLT
853	853	DROP
854	24	AUTOMATI
	166	MAGNETIC
	404	MACHINE
	441	READING
	484	ACCESS
	511	COMPUT
	687	DIGIT
	854	DRUM
	919	READ
	938	STOR
	940	TAPE
855	499	BRIDGE
	774	RELAY
	855	DUAL
856	231	PROPAGAT
	651	WHISTL
	655	ALIGN
	856	DUCT
857	115	GEOPHYSI
	736	MAJOR
	828	AREA
	857	EAST
858	157	IRREGULA
	269	SIMULTAN
	272	SOUNDING
	280	SPORADIC
	494	ASPECT
	567	METEOR
	619	SPREAD
	649	VISUAL
	769	RADAR
	810	TRAIL
	858	ECHO
859	818	WEDGE
	859	EDGE
860	531	EMISSI
	860	EMIT
861	861	FACE
862	862	FAIL
863	240	RECOMBIN
	481	VIRTUAL
	519	DENSIT
	580	OXYGEN
	751	NIGHT
	863	FALL
864	589	PRESET
	610	SELECT
	636	TIMING
	864	FAST
865	59	CONVENTI
	147	INTERNAL
	724	INPUT
	827	ANOD
	852	DRIV
865	865	FEED
	875	GRID
866	65	CRYOTRON
	92	EVAPORAT
	365	DEPOSIT
	501	CARBON
	866	FILM
	941	THIN
867	23	ATTENUAT
	138	INSERTIO
	198	NORMALIZ
	474	TORSION
	498	BRANCH
	556	LADDER
	719	IMAGE
	738	MATCH
	831	BAND
	867	FILT
	905	PASS
868	288	STRUCTUR
	667	BURST
	868	FINE
869	32	BISTABLE
	49	COMPLEME
	183	MONOSTAB
	190	MULTIVIB
	201	OPERATIO
	314	TRANSIST
	477	TRIGGER
	630	SWITCH
	869	FLIP
870	708	FORCE
	870	FLOW
871	111	GALACTIC
	235	RADIATIO
	515	COSMIC
	871	FLUX
872	872	FORM
873	873	FREE
874	29	BANDWIDT
	59	CONVENTI
	491	AMPLIF
	642	TUNNEL
	704	FIGUR
	838	CHOP
	874	GAIN
	914	PULL
	915	PUMP
	935	STAB
	997	DB
875	423	PENTODE
	461	STABILI
	639	TRIODE
	814	VALVE
	827	ANOD
	835	BIAS
	865	FEED
	875	GRID
	914	PULL
876	44	COLLISIO
	100	EXTRAORD
	406	MAGNETO
	876	GYRO
877	166	MAGNETIC
	597	RECORD
	877	HEAD
	919	READ
	940	TAPE
	959	WRIT
878	472	THERMAL
	878	HEAT
879	34	BREMSSTR
	253	RESOLUTI
	713	GLASS
	800	SPEED
	879	HIGH
	986	OIL
880	18	ASSOCIAT
	846	CURV
	880	HISS
881	881	HOLE
882	324	VARIATIO
	372	DIURNAL
	728	LAYER
	882	HOUR
	969	DAY
883	213	PHENOMEN
	257	RESPONSE
	658	ANALY
	883	JUMP
884	884	LAND
885	885	LEAD
886	886	LEAK
887	887	LIFE
888	465	STRAIGH
	657	AMMON
	717	HELIC
	812	TWIST
	840	COAX
	888	LINE
889	791	SHUNT
	889	LOAD
890	108	FREQUENC
	203	OSCILLAT
	588	PRECIS
	589	PRESET
	849	DIVI
	890	LOCK
	908	PHAS
891	891	LONG
	939	TAIL
892	101	FEEDBACK
	507	CLOSED
	782	SERVO
	892	LOOP
893	23	ATTENUAT
	138	INSERTIO
	893	LOSS
	971	EDD
894	114	GEOMAGNE
	529	DYNAMO
	693	EARTH
	803	STORM
	894	MAIN
	939	TAIL
	996	ZON
895	895	MASK
896	669	CHARG
	896	MASS
897	749	MONTH
	897	MEAN
898	281	SPURIOUS
	648	VIBRAT
	785	SHEAR
	812	TWIST

898	898	MODE
899	899	MOON
	923	ROCK
	983	LUN
900	26	BACKGROU
	292	SUPERREG
	704	FIGUR
	820	WHITE
	900	NOIS
	915	PUMP
	927	SHOT
901	158	IRREVERS
	901	NOTE
902	902	OPEN
903	92	EVAPORAT
	262	SECONDAR
	685	DEPTH
	743	METAL
	839	COAT
	903	OXID
904	239	RECIPROC
	904	PAIR
905	19	ASYMMETR
	138	INSERTIO
	263	SELECTIV
	299	TERMINAT
	462	STAGGER
	498	BRANCH
	502	CASCAD
	556	LADDER
	719	IMAGE
	738	MATCH
	831	BAND
	867	FILT
	905	PASS
	982	LOW
906	906	PATH
907	645	VALLEY
	907	PEAK
908	265	SEMIDIUR
	552	INVERT
	788	SHIFT
	890	LOCK
	908	PHAS
909	216	PHOTOELE
	829	ATOM
	909	PHOT
910	257	RESPONSE
	910	PLOT
911	109	FUNCTION
	311	TRANSFER
	911	POLE
912	912	PORT
913	913	PROD
	988	RED
914	28	BALANCED
	437	QUALITY
	671	CLASS
	874	GAIN
	875	GRID
	914	PULL
915	29	BANDWIDT
	222	POPULATI
	419	PARAMET
	491	AMPLIF
	704	FIGUR
	737	MASER
915	874	GAIN
	900	NOIS
	915	PUMP
	978	IDL
916	720	IMPUR
	916	PURE
917	478	TUNABLE
	917	RANG
918	918	RATE
919	441	READING
	511	COMPUT
	800	SPEED
	854	DRUM
	877	HEAD
	919	READ
	938	STOR
	940	TAPE
	959	WRIT
	966	BIT
920	109	FUNCTION
	221	POLYNOMI
	920	REAL
921	770	RADII
	921	RING
922	33	BLOCKING
	180	MILLIMIC
	388	GENERAT
	417	OVERSHO
	766	PULSE
	922	RISE
	942	TIME
923	469	SURFACE
	899	MOON
	923	ROCK
924	296	TEMPERAT
	924	ROOM
925	13	APPARATU
	253	RESOLUTI
	295	TELESCOP
	925	SCAN
926	203	OSCILLAT
	926	SELF
927	105	FLUCTUAT
	900	NOIS
	927	SHOT
928	928	SIGN
929	929	SIZE
930	350	CONDUCT
	492	ANOMAL
	685	DEPTH
	743	METAL
	930	SKIN
931	40	CIRCULAR
	72	DIFFRACT
	391	INCIDEN
	607	SCREEN
	759	PLANE
	804	STRIP
	931	SLIT
	954	WAVE
	972	FAR
932	932	SLOW
933	106	FORMATIO
	933	SOFT
	987	RAY
	999	X
934	103	FERROMAG
	167	MAGNETIS
	193	NONMAGNE
	286	STOCHAST
	569	MOMENT
	702	FERRI
	795	SOLID
	934	SPIN
935	5	ADVANTAG
	314	TRANSIST
	461	STABILI
	535	EXTREM
	639	TRIODE
	642	TUNNEL
	874	GAIN
	935	STAB
936	261	SCINTILL
	771	RADIO
	936	STAR
937	257	RESPONSE
	313	TRANSIEN
	417	OVERSHO
	937	STEP
938	65	CRYOTRON
	136	INFORMAT
	404	MACHINE
	484	ACCESS
	742	MEMOR
	800	SPEED
	854	DRUM
	919	READ
	938	STOR
	940	TAPE
	966	BIT
939	529	DYNAMO
	747	MODEL
	847	DARK
	891	LONG
	894	MAIN
	939	TAIL
940	136	INFORMAT
	364	DECIMAL
	441	READING
	511	COMPUT
	597	RECORD
	687	DIGIT
	800	SPEED
	854	DRUM
	877	HEAD
	919	READ
	938	STOR
	940	TAPE
941	743	METAL
	866	FILM
	941	THIN
942	150	INTERVAL
	417	OVERSHO
	443	RECOVER
	475	TRANSIT
	922	RISE
	942	TIME
	980	LAG
943	108	FREQUENC
	293	SYNCHRON
	943	TONE
944	208	PARTICLE
	366	DETONAT
	590	PROTON
	834	BELT
	944	TRAP
	987	RAY
945	945	TUBE

946	946	UNIF
947	947	UNIT
	962	ADD
948	92	EVAPORAT
	743	METAL
	948	VACU
949	949	VARY
950	308	TOLERANC
	429	PLASTIC
	761	POINT
	825	ADHE
	836	BOND
	950	VIEW
951	83	EFFICIEN
	307	THYRATRO
	440	REACTOR
	446	REGULAT
	468	SUPPLIE
	497	BALANC
	543	HEATER
	781	SELEN
	852	DRIV
	951	VOLT
	986	OIL
	995	TUN
	998	DC
952	952	WALL
953	953	WASH
954	1	ACOUSTIC
	27	BACKWARD
	63	CORRUGAT
	72	DIFFRACT
	86	ELECTROM
	99	EXTINCTI
	179	MILLIMET
	231	PROPAGAT
	283	STANDING
	289	SUBMILLI
	318	TRAVELLI
	565	MEDIUM
	607	SCREEN
	741	MEDIA
	931	SLIT
	954	WAVE
955	955	WELL
956	21	ATMOSPHE
	122	HORIZONT
	537	FADING
	956	WIND
957	823	WOUND
	957	WIRE
958	958	WORK
959	441	READING
	877	HEAD
	919	READ
	959	WRIT
960	2	ACTIVITY
	115	GEOPHYSI
	148	INTERNAT
	162	LATITUDE
	324	VARIATIO
	463	STATION
	467	SUNSPOT
	564	MEDIAN
	609	SEASON
	678	CYCLE
	794	SOLAR
	960	YEAR
961	961	ZERO
962	201	OPERATIO
	364	DECIMAL
	579	OUTPUT
	665	BINAR
	675	COUNT
	687	DIGIT
	800	SPEED
	947	UNIT
	962	ADD
963	21	ATMOSPHE
	850	DRAG
	963	AIR
964	111	GALACTIC
	500	BRIGHT
	573	NEBULA
	964	ARC
	988	RED
965	114	GEOMAGNE
	178	MIDNIGHT
	496	AURORA
	591	PULSAT
	752	NORTH
	803	STORM
	965	BAY
966	364	DECIMAL
	484	ACCESS
	589	PRESET
	632	SYSTEM
	687	DIGIT
	735	LOGIC
	742	MEMOR
	919	READ
	938	STOR
	966	BIT
967	40	CIRCULAR
	72	DIFFRACT
	121	HOMOGENE
	278	SPHEROID
	395	INFINIT
	513	CONVEX
	594	RADIUS
	796	SOLVE
	967	BOD
968	364	DECIMAL
	441	READING
	687	DIGIT
	724	INPUT
	968	COD
969	627	SUMMER
	652	WINTER
	678	CYCLE
	768	QUIET
	882	HOUR
	969	DAY
970	970	DIS
	993	SUN
971	361	CURRENT
	893	LOSS
	971	EDD
972	931	SLIT
	972	FAR
973	252	RESISTOR
	323	VARIABLE
	341	CAPACIT
	973	FIX
974	974	FOC
975	975	GAP
976	316	TRANSPOR
976	413	NEUTRAL
	621	STATIC
	710	GASES
	712	GIANT
	726	IONIZ
	765	PROBE
	976	GAS
977	30	BIDIRECT
	589	PRESET
	630	SWITCH
	665	BINAR
	977	GAT
978	108	FREQUENC
	419	PARAMET
	491	AMPLIF
	613	SIGNAL
	642	TUNNEL
	704	FIGUR
	915	PUMP
	978	IDL
	990	ROW
979	52	CONCENTR
	87	ELECTRON
	191	NITROGEN
	413	NEUTRAL
	524	DIFFUS
	710	GASES
	765	PROBE
	979	ION
980	942	TIME
	980	LAG
981	472	THERMAL
	981	LAW
982	535	EXTREM
	704	FIGUR
	905	PASS
	982	LOW
983	14	APPARENT
	265	SEMIDIUR
	325	VERTICAL
	899	MOON
	983	LUN
	994	TID
984	108	FREQUENC
	112	GAUSSIAN
	613	SIGNAL
	984	MIX
985	84	ELECTRIC
	985	NET
986	879	HIGH
	951	VOLT
	986	OIL
987	34	BREMSSTR
	149	INTERPLA
	300	TERRESTR
	335	BALLOON
	414	NEUTRON
	515	COSMIC
	590	PROTON
	705	FLARE
	834	BELT
	933	SOFT
	944	TRAP
	987	RAY
988	235	RADIATIO
	913	PROD
	964	ARC
	988	RED
989	989	ROD
990	192	NONLINEA

990	237	REACTANC
	256	RESTRICT
	439	REACTIV
	763	POWER
	978	IDL
	990	ROW
991	991	SET
992	992	SUM
993	145	INTERFER
	235	RADIATIO
	300	TERRESTR
	500	BRIGHT
	514	CORONA
	661	ASTRO
	794	SOLAR
	828	AREA
	970	DIS
993	993	SUN
994	184	MORPHOLO
	265	SEMIDIUR
	324	VARIATIO
	529	DYNAMO
	983	LUN
	994	TID
995	446	REGULAT
	462	STAGGER
	468	SUPPLIE
	814	VALVE
	951	VOLT
	995	TUN
	998	DC
996	496	AURORA
	762	POLAR
	894	MAIN
996	996	ZON
997	23	ATTENUAT
	874	GAIN
	997	DB
998	446	REGULAT
	468	SUPPLIE
	497	BALANC
	763	POWER
	814	VALVE
	838	CHOP
	951	VOLT
	995	TUN
	998	DC
999	34	BREMSSTR
	826	ALTI
	933	SOFT
	999	X

APPENDIX A6

SELECTION FROM CLUSTER SET C3, USED IN RUN 14(MCS01)

...
...

6 AFTERNOO
156 IONOSPHE
196 NONSEASO
324 VARIATIO
362 DAYTIME
372 DIURNAL
609 SEASON
627 SUMMER
652 WINTER
678 CYCLE
679 DAILY
728 LAYER

6 AFTERNOO
156 IONOSPHE
196 NONSEASO
324 VARIATIO
362 DAYTIME
372 DIURNAL
609 SEASON
627 SUMMER
652 WINTER
678 CYCLE
679 DAILY
960 YEAR

6 AFTERNOO
178 MIDNIGHT
196 NONSEASO
372 DIURNAL
728 LAYER

6 AFTERNOO
178 MIDNIGHT
372 DIURNAL
733 LOCAL

6 AFTERNOO
196 NONSEASO
324 VARIATIO
372 DIURNAL
564 MEDIAN
609 SEASON
652 WINTER
678 CYCLE
679 DAILY
728 LAYER

6 AFTERNOO
196 NONSEASO
324 VARIATIO
372 DIURNAL
564 MEDIAN
609 SEASON
652 WINTER
678 CYCLE
679 DAILY
960 YEAR

6 AFTERNOO
196 NONSEASO
324 VARIATIO
372 DIURNAL
609 SEASON
627 SUMMER
652 WINTER
679 DAILY
728 LAYER
882 HOUR

6 AFTERNOO
196 NONSEASO
466 SUNRISE
609 SEASON
679 DAILY
728 LAYER

6 AFTERNOO
564 MEDIAN
609 SEASON
652 WINTER
678 CYCLE
679 DAILY
749 MONTH
960 YEAR

...
...

17 ARTIFICI
21 ATMOSPHE
260 SATELLIT
416 ORBITAL
424 PERIGEE
693 EARTH
755 ORBIT
850 DRAG

21 ATMOSPHE
159 ISOTHERM

21 ATMOSPHE
163 LIGHTNIN
672 CLOUD

21 ATMOSPHE
236 RADIOSON

21 ATMOSPHE
416 ORBITAL
850 DRAG
963 AIR

21 ATMOSPHE
580 OXYGEN

21 ATMOSPHE
813 UPPER

21 ATMOSPHE
956 WIND

...
...

24 AUTOMATI
136 INFORMAT
363 DECIMAL
441 READING
632 SYSTEM
687 DIGIT
854 DRUM
966 BIT

24 AUTOMATI
136 INFORMAT
363 DECIMAL
441 READING
665 BINAR
687 DIGIT
854 DRUM
940 TAPE
966 BIT
968 COD

24 AUTOMATI
136 INFORMAT
363 DECIMAL
441 READING
687 DIGIT
854 DRUM
919 READ
940 TAPE
966 BIT

24 AUTOMATI
136 INFORMAT
404 MACHINE
632 SYSTEM
687 DIGIT
854 DRUM
966 BIT

24 AUTOMATI
136 INFORMAT
404 MACHINE
687 DIGIT
854 DRUM
919 READ
940 TAPE
966 BIT

...
...

44 COLLISIO
87 ELECTRON
316 TRANSPOR
401 KINETIC
621 STATIC

44 COLLISIO
87 ELECTRON
316 TRANSPOR
413 NEUTRAL
524 DIFFUS
726 IONIZ
976 GAS

44 COLLISIO
87 ELECTRON
316 TRANSPOR
413 NEUTRAL
621 STATIC
726 IONIZ
976 GAS

44 COLLISIO
87 ELECTRON
413 NEUTRAL
524 DIFFUS
726 IONIZ
976 GAS
979 ION

44 COLLISIO
87 ELECTRON
413 NEUTRAL
587 PLASMA
621 STATIC
726 IONIZ
976 GAS

44 COLLISIO
87 ELECTRON
413 NEUTRAL
587 PLASMA
726 IONIZ
976 GAS
979 ION

44 COLLISIO
100 EXTRAORD
406 MAGNETO
876 GYRO

44 COLLISIO
121 HOMOGENE
246 RELATIVI
316 TRANSPOR
401 KINETIC
621 STATIC

44 COLLISIO
121 HOMOGENE
246 RELATIVI
316 TRANSPOR
401 KINETIC
708 FORCE

44 COLLISIO
121 HOMOGENE
246 RELATIVI
316 TRANSPOR
413 NEUTRAL
621 STATIC

44 COLLISIO
121 HOMOGENE
316 TRANSPOR
401 KINETIC
621 STATIC
703 FIELD

44 COLLISIO
121 HOMOGENE
316 TRANSPOR
401 KINETIC
703 FIELD
708 FORCE

44 COLLISIO
123 HYDRODYN
316 TRANSPOR
406 MAGNETO
413 NEUTRAL
725 IONIC
726 IONIZ

44 COLLISIO
123 HYDRODYN
406 MAGNETO
413 NEUTRAL
587 PLASMA
725 IONIC
726 IONIZ

44 COLLISIO
123 HYDRODYN
406 MAGNETO
725 IONIC
876 GYRO

44 COLLISIO
131 INCOHERE
246 RELATIVI
413 NEUTRAL
621 STATIC

44 COLLISIO
131 INCOHERE
413 NEUTRAL
587 PLASMA
621 STATIC
976 GAS

...
...

APPENDIX B1

COMPLETE LIST OF REQUESTS

showing the name of the requestor, the number of key-word stems, the stems themselves with their underlinings, if any, (section VI.2.6), and their dictionary code numbers (Appendix A1).

Request 0 (P.H. Hammond) (6)
HIGH FREQUENCY OSCILLATORS USING TRANSISTORS THEORETICAL TREATMENT AND PRACTICAL CIRCUIT DETAILS
HIGH FREQUENC OSCILLAT <u>TRANSIST</u> THEOR CIRCUIT
879 108 203 314 807 345

Request 1 (Cancelled) (10)
SUPER REGENERATIVE RECEIVERS IN THE HIGH FREQUENCY REGION WITH DETAILS OF NOISE FIGURE SELECTIVITY AND RADIATED
REGENERA RECEIV HIGH FREQUENC NOIS FIGURE SELECTIV RADIAT
244 596 879 108 900 704 263 593
SIGNAL LEVEL
SIGNAL LEVEL
613 730

Request 2 (P.H. Hammond) (5)
MEASUREMENT OF DIELECTRIC CONSTANT OF LIQUIDS BY THE USE OF MICROWAVE TECHNIQUES
MEASUR <u>DIELECTR</u> CONSTAN LIQUID MICROWAV
563 70 352 560 177

Request 3 (P.H. Hammond) (6)
MATHEMATICAL ANALYSIS AND DESIGN DETAILS OF WAVEGUIDE FED MICROWAVE RADIATIONS
MATHEMAT ANALY DESIGN WAVEGUID MICROWAV RADIAT
172 658 520 327 177 593

Request 4 (P.H. Hammond) (9)
USE OF DIGITAL COMPUTERS IN THE DESIGN OF BAND PASS FILTERS HAVING GIVEN PHASE AND ATTENUATION CHARACTERISTICS
DIGIT <u>COMPUT</u> DESIGN BAND PASS FILT PHAS ATTENUAT CHARACTE
687 512 520 831 905 867 908 23 39

Request 5 (F.M. Blake) (5)
SYSTEMS OF DATA CODING FOR INFORMATION TRANSFER
SYSTEM DATA COD INFORMAT TRANSFER
632 848 968 136 311

Request 6 (Cancelled) (3)
USE OF PROGRAMS IN ENGINEERING TESTING OF COMPUTERS
PROGRAM ENGINEER COMPUT
435 89 512

Request 7 (F.M. Blake) (5)
SIMULATION OF MATHEMATICAL FUNCTIONS USING MAGNETIC CIRCUITS
SIMULT MATHEMAT FUNCTION <u>MAGNETIC</u> CIRCUIT
458 172 109 166 345

Request 8 (Cancelled) (4)
MULTIPLE DIGIT TECHNIQUES IN FONT DECIMAL ADDRESS
MULTIPL DIGIT DECIMAL ADD
409 687 363 962

Request 9 (F.M. Blake) (4)
NUMBER REPRESENTATION IN BINARY MACHINES
NUMBER REPRESEN BINAR MACHINE
575 248 665 404

Request 10 (W.L. Price) (7)
SECONDARY EMISSION OF ELECTRONS BY POSITIVE ION BOMBARDMENT OF THE CATHODE
<u>SECONDAR</u> <u>EMISSI</u> <u>ELECTRON</u> <u>POSITIVE</u> <u>ION</u> <u>BOMBARD</u> CATHOD
262 531 87 224 979 338 503

Request 11 (W.L. Price) (7)
MEASUREMENT OF PLASMA TEMPERATURES IN ARC DISCHARGE USING SHOCK WAVE TECHNIQUES
MEASUR PLASMA <u>TEMPERAT</u> <u>ARC</u> DISCHARG <u>SHOCK</u> <u>WAVE</u>
563 587 296 964 76 789 954

Request 12 (W.L. Price) (7)
CHARACTERISTICS OF THE SINGLE ELECTRODE DISCHARGE IN THE RARE GASES AT LOW PRESSURES
CHARACTE <u>SINGL</u> <u>ELECTROD</u> DISCHARG GASES LOW PRESSURE
39 793 85 76 710 982 227

Request 13 (W.L. Price) (5)
METHODS OF CALCULATING INSTANTANEOUS POWER DISSIPATION IN REACTIVE CIRCUITS
CALCULAT POWER DISSIPA REACTIV CIRCUIT
35 763 79 439 345

Request 14 (W.L. Price) (3)
THE EFFECT OF OXIDATION ON CIRCUIT BREAKER CONTACTS
<u>OXID</u> CIRCUIT <u>CONTACT</u>
903 345 353

Request 15 (W. Fincham) (7)
TEMPERATURE INDEPENDENT METHODS FOR TUNING HIGHLY STABLE HIGH FREQUENCY OSCILLATORS
TEMPERAT INDEPEND TUN (HIGH) STAB <u>HIGH</u> <u>FREQUENC</u> OSCILLAT
296 133 995 (879) 935 879 108 203

Request 16 (W. Fincham) (8)
MATHEMATICAL EXPRESSIONS AND GRAPHS FOR THE DESIGN OF TRANSISTORISED TUNED PASS BAND AMPLIFIERS
MATHEMAT GRAPH DESIGN TRANSIST TUN PASS BAND AMPLIF
172 714 520 314 995 905 831 491

Request 17 (W. Fincham) (6)
SYNTHESIS OF FILTERS HAVING HIGH ATTENUATION CHARACTERISTICS USING A BLOCK DIAGRAM APPROACH
SYNTHESI FILT HI ATTENUAT CHARACTE BLOCK
294 867 879 23 39 666

Request 18 (W. Fincham) (7)
VOLTAGE CURRENT RELATIONSHIPS IN NETWORKS OF NONLINEAR ELEMENTS CONNECTED IN PARALLEL
VOLT CURRENT NETWORK NONLINEAR ELEMENT CONNE PARALLEL
954 361 412 192 376 674 206

Request 19 (W. Fincham) (7)
METHODS OF APPROXIMATING THE FREQUENCY PHASE RELATIONSHIPS FOR RESISTIVE INDUCTIVE AND RESISTIVE CAPACITIVE
APPROXIM FREQUENC PHAS RESISTIV INDUCTI (RESISTIV) CAPACIT
15 108 908 251 393 251 341

CIRCUITS
CIRCUIT
345

Request 20 (D.C. Bain) (9)
OBSERVATIONS OF RAPID FLUCTUATIONS IN THE EARTHS MAGNETIC FIELD AND THEIR RELATION TO THE PROPAGATION OF
OBSERVAT FLUCTUAT EARTH MAGNETIC FIELD PROPAGAT
200 105 693 166 703 231

HYDROMAGNETIC WAVES IN THE EXOSPHERE
HYDROMAG WAVE EXOSPHER
125 954 95

Request 21 (D.C. Bain) (6)
DIURNAL VARIATIONS OF FLUCTUATIONS IN THE EARTHS MAGNETIC FIELD
DIURNAL VARIATIO FLUCTUAT EARTH MAGNETIC FIELD
372 324 105 693 166 703

Request 22 (E. Dunford) (6)
SPHERICAL HARMONIC ANALYSIS OF THE EARTHS MAGNETIC FIELD
SPHERICA HARMONIC ANALY EARTH MAGNETIC FIELD
277 199 658 693 166 703

Request 23 (E. Dunford) (6)
DERIVATION OF THE COMPONENTS OF THE ELECTRICAL CONDUCTIVITY IN THE UPPER ATMOSPHERE
DERIVAT COMPONEN ELECTRIC CONDUCT UPPER ATMOSPHE
69 50 84 350 813 21

Request 24 (H. Rishbeth) (7)
THE EFFECTS OF SOLAR FLARES ON THE ABSORPTION OF COSMIC RADIO NOISE IN THE IONOSPHERE
SOLAR FLARE ABSOR COSMIC RADIO NOIS IONOSPHE
794 705 653 515 771 900 156

Request 25 (E. Dunford) (6)
DETERMINATION OF THE CURRENT SYSTEMS IN THE UPPER ATMOSPHERE DURING MAGNETIC STORMS
CURRENT SYSTEM UPPER ATMOSPHE MAGNETIC STORM
361 632 813 21 166 803

Request 26 (D.L. Croom) (8)
OBSERVATIONS OF THE SUN DURING ECLIPSES GIVING THE DISTRIBUTION OF SOURCES ON THE DISC IN THE MICROWAVE RANGE
OBSERVAT SUN ECLIPS DISTRIBU SOURCE DIS MICROWAV RANG
200 993 530 80 615 970 177 917

Request 27 (D.L. Croom) (4)
OBSERVATIONS OF THE SUN USING RADIO INTERFEROMETERS
OBSERVAT SUN RADIO INTERFER
200 993 771 145

Request 28 (D.L. Croom) (7)
EQUATIONS GOVERNING THE PROPAGATION OF ELECTROMAGNETIC AND HYDROMAGNETIC WAVES IN THE SOLAR CORONA
EQUAT PROPAGAT ELECTROM HYDROMAG WAVE SOLAR CORONA
696 231 86 125 954 794 514

Request 29 (D.L. Croom) (
ESTIMATES OF THE DENSITY OF IONIZATION AND TEMPERATURE IN THE SOLAR CORONA
DENSIT IONIZATI TEMPERAT SOLAR CORONA
519 153 296 794 514

Request 30 (D.L. Croom) (4)
PLEASE SEND ABSTRACTS ON THE SOURCE SPECTRA OF LIGHTNING DISCHARGES
SOURCE SPECTR LIGHTNIN DISCHARG
616 616 163 76

Request 31 (Cancelled) (5)
PLEASE SEND ABSTRACTS CONNECTED WITH THE OCCURRENCE OF A SECONDARY OZONE LAYER AT NIGHT TIME
CONNE SECONDAR LAYER NIGHT TIME
674 262 728 751 942

Request 32 (D.M. Yates) (4)
THE EFFECT OF SMALL DISTORTIONS IN THE SURFACE OF A CAVITY RESONATOR
DISTORT SURFACE CAVIT RESONATO
370 469 668 255

Request 33 (W.C. Bain) (3)
THE DETERMINATION OF THE ORBITS OF INDIVIDUAL METEORS BY RADIO METHODS
ORBIT METEOR RADIO
755 567 771

Request 34 (W.C. Bain) (5)
THE DETERMINATION OF ION MASSES IN THE IONOSPHERE BY THE STUDY OF BACK SCATTERED RADIO WAVES
ION IONOSPHE SCATTER RADIO WAVE
979 156 452 771 954

Request 35 (W.C. Bain) (9)
THEORETICAL STUDIES OF THE SOURCE OF HIGH FREQUENCY RADIO WAVES EMITTED FROM THE PLANET JUPITER
THEOR SOURCE HIGH FREQUENC RADIO WAVE EMIT PLANET JUPITER
807 615 879 108 771 954 860 586 400

Request 36 (W.C. Bain) (5)
SIMULTANEOUS OBSERVATIONS OF WHISTLERS AND LIGHTNING DISCHARGES
SIMULTAN OBSERVAT WHISTL LIGHTNIN DISCHARG
269 200 651 163 76

Request 37 (W.C. Bain) (10)
VARIATIONS IN THE HEIGHT OF REFLECTION OF LOW OR VERY LOW FREQUENCY RADIO WAVES IN THE PERIOD BEFORE GROUND
VARIATI HEIGHT REFLECT LOW (LOW) FREQUENC RADIO WAVE PERIOD GROUND
324 544 444 982 (982) 108 771 954 582 541

SUNRISE
SUNRISE
466

Request 38 (H. Rishbeth) (7)
THE USE OF IONOSPHERIC CROSS MODULATION IN THE DETERMINATION OF IONOSPHERIC ELECTRON DENSITIES AND COLLISION
CROSS MODULAT IONOSPHE ELECTRON DENSIT COLLISIO
677 407 156 87 519 44

FREQUENCIES
FREQUENC
108

Request 39 (H. Rishbeth) (4)
MEASUREMENTS OF IONOSPHERIC DRIFTS NEAR THE EQUATOR
MEASUR IONOSPHE DRIFT EQUATOR
563 156 692 379

Request 40 (H. Rishbeth) (5)
THE USE OF ANALOGUE COMPUTERS IN UPPER ATMOSPHERE THEORY
ANALOGUE COMPUT UPPER ATMOSPHE THEOR
10 512 813 21 807

Request 41 (H. Rishbeth) (5)
THE USE OF LUNAR RADIO REFLECTIONS IN INVESTIGATIONS OF THE NATURE OF THE MOONS SURFACE
LUN RADIO REFLECT MOON SURFACE
983 771 444 899 469

Request 42 (D.A. Bryant) (9)
WHAT REFERENCES ON COMPARISON BETWEEN GROUND LEVEL AND HIGH ALTITUDE BALLOON COSMIC RAY RESULTS
REFERENC COMPARIS GROUND LEVEL HIGH ALTI BALLOON COSMIC RAY
241 46 541 730 879 826 335 515 987

Request 43 (D.A. Bryant) (3)
WHAT REFERENCES ON MODEL EXPERIMENTS ON AURORA
REFERENC MODEL AURORA
241 747 496

Request 44 (D.A. Bryant) (6)
WHAT REFERENCES ON INTEGRAL SPECTRUM OF PRIMARY COSMIC RAYS
REFERENC INTEGRAL SPECTR PRIMARY COSMIC RAY
241 141 616 432 515 987

Request 45 (Cancelled) (5)
REFERENCES ON SOLAR AND GEOMAGNETIC EFFECTS OF COSMIC RAYS
REFERENC SOLAR GEOMAGNE COSMIC RAY
[illegible] 794 114 515 987

Request 46 (Cancelled) (4)
CODING DISCS FOR ANALOGUE DIGITAL CONVERTERS
COD ANALOGUE DIGIT CONVERT
968 10 687 357

Request 47 (J. McDaniel) (3)
FERROMAGNETIC TECHNIQUES FOR COMPUTER STORES
FERROMAG COMPUT STOR
103 512 938

Request 48 (J. McDaniel) (4)
SOLUTION OF DIFFERENTIAL EQUATIONS BY COMPUTER
SOLUTION DIFFEREN EQUAT COMPUT
271 71 696 512

Request 49 (J. McDaniel) (6)
EFFICIENCY OF DIGITAL COMPUTERS VERSUS ANALOGUE COMPUTERS IN THE SOLUTION OF BOUNDARY VALUE PROBLEMS
EFFICEN DIGIT ANALOGUE COMPUT SOLUTION BOUNDAR
83 687 10 512 271 339

Request 50 (J. McDaniel) (3)
METHODS OF ERROR CHECKING IN DIGITAL COMPUTERS
ERROR DIGIT COMPUT
697 687 512

Request 51 (M. Longden) (7)
METHODS OF PRODUCING MINIMAL NETS GIVEN A LOGICAL FUNCTION IN CANONICAL FORM
PROD MINIM NET LOGIC FUNCTION CANONIC FORM
913 745 985 735 109 340 872

Request 52 (M. Longden) (6)
ARITHMETIC UNITS AS REQUIRED IN A DIGITAL COMPUTER INCLUDING SHIFT REGISTERS SERIAL AND PARALLEL ADDER
UNIT DIGIT COMPUT SHIFT PARALLEL ADD
947 687 512 788 206 962

Request 53 (C.H. Davis) (6)
INFORMATION ON THE DESIGN OF LOW DRIFT TRANSISTOR AMPLIFIERS
INFORMAT DESIGN LOW DRIFT TRANSIST AMPLIF
136 520 982 692 314 491

Request 54 (C.H. Davis) (5)
INFORMATION ON HIGH CURRENT TRANSISTOR SWITCHES
INFORMAT HIGH CURRENT TRANSIST SWITCH
136 879 361 314 630

Request 55 (C.H. Davis) (6)
INFORMATION ON DESIGN OF TIME DIVISION MULTIPLEXING CIRCUITS
INFORMAT DESIGN TIME DIVI MULTIPL CIRCUIT
136 520 942 849 409 345

Request 56 (C.H. Davis) (3)
DETAILS OF AVAILABLE LOW VOLTAGE CAPACITORS
LOW VOLT CAPACIT
982 951 341

Request 57 (C.H. Davis) (9)
DESIGN OF DIRECT COUPLED FLIP FLOPS TO FUNCTION WITH THE MAXIMUM VARIATIONS IN THE VALUES OF THE CIRCUIT
DESIGN DIRECT COUPL FLIP FUNCTION MAXIM VARIATIO CIRCUIT
520 525 676 869 109 740 324 345

COMPONENTS
COMPONEN
50

Request 58 (A.A. Hill) (7)
PLEASE SUPPLY INFORMATION ON THE PERFORMANCE OF TYPICAL MAGNETIC FILM MEMORY SYSTEMS WITH CIRCUIT DIAGRAMS
INFORMAT PERFORMA MAGNETIC FILM MEMOR SYSTEM CIRCUIT
136 210 166 866 742 632 345

Request 59 (A.A. Hill) (6)
I WOULD LIKE DETAILS OF THE WORK WHICH HAS BEEN DONE TO EXTEND THE FREQUENCY RANGE OF MAGNETIC AMPLIFIERS
WORK EXTEND FREQUENC RANG MAGNETIC AMPLIF
6 534 108 917 166 491

Request 60 (A.A. Hill) (6)
I WOULD LIKE INFORMATION ON THE RANGE OF STATIC RELAYS SUITABLE FOR USE AT HIGH SWITCHING RATES
INFORMAT RANG STATIC RELAY HIGH SWITCH
136 917 621 774 879 630

Request 61 (A.A. Hill) (5)
I AM INTERESTED IN CIRCUITRY CAPABLE OF GENERATING EXTREMELY NARROW PULSES
CIRCUIT GENERAT EXTREM NARROW PULSE
345 388 535 572 766

Request 62 (A.A. Hill) (4)
PLEASE SUPPLY INFORMATION ON THE THEORY AND USE OF PARAMETRIC AMPLIFIERS
INFORMAT THEOR PARAMET AMPLIF
136 807 419 491

Request 63 (P.A. Briggs) (6)
THE SYNTHESIS OF NETWORKS WITH GIVEN SAMPLED DATA TRANSFER FUNCTIONS
SYNTHESI NETWORK SAMPL DATA TRANSFER FUNCTION
294 412 779 848 311 109

Request 64 (P.A. Briggs) (6)
THE THEORY OF STABILITY IN RELATION TO CONTROL SYSTEMS WITH RELAY ELEMENTS
THEOR STABILI CONTROL SYSTEM RELAY ELEMENT
807 461 356 632 774 376

Request 65 (P.A. Briggs) (7)
THE USE OF DIGITAL COMPUTERS TO OBTAIN POWER SPECTRAL ANALYSIS OF NUMERICAL DATA
DIGIT COMPUT POWER SPECTR ANALY NUMERICA DATA
687 512 763 616 658 199 848

Request 66 (Cancelled) (8)
CIRCUIT ELEMENTS WITH ZERO MEMORY NON LINEAR CHARACTERISTICS AND VARIABLE PARAMETERS
CIRCUIT ELEMENT ZERO MEMOR LINEAR CHARACTE VARIABLE PARAMET
345 376 961 742 559 39 323 419

Request 67 (P.A. Briggs) (8)
PRINTED CIRCUIT DESIGN FOR A RANDOM PULSE GENERATOR OF LOW FREQUENCY
PRINT CIRCUIT DESIGN RANDOM PULSE GENERAT LOW FREQUENC
764 108 520 595 768 388 982 108

Request 68 (D.A. Bryant) (5)
REFERENCES ON ELECTRIC FIELD THEORIES OF THE AURORA
REFERENC ELECTRIC FIELD THEOR AURORA
241 34 703 807 496

Request 69 (C.H.Davis) (3)
FAST TRANSISTOR COUNTERS
FAST TRANSIST COUNT
864 314 675

Request 70 (E.A. Newman) (4)
LOW PASS LATTICE FILTERS
LOW PAS LATTICE FILT
982 905 402 867

Request 71 (D.M. Yates) (7)
SIMILARITIES BETWEEN THE DIFFRACTION THEORY OF ELECTROMAGNETIC WAVES AND THAT OF ELECTRON STREAMS
SIMILAR DIFFRACT THEOR ELECTROM WAVE ELECTRON STREAM
457 72 807 86 954 87 624

Request 72 (D.M. Yates) (5)
THE USE OF COMPLEX VARIABLES IN THE THEORY OF COMMUNICATION NETWORKS
COMPLEX VARIABLE THEOR COMMUNIC NETWORK
349 323 807 45 412

Request 73 (D.M. Yates) (5)
THE BEHAVIOUR OF A BEAM OF CHARGED PARTICLES IN THE PRESENCE OF PLANE CONDUCTORS
BEAM CHARG PARTICLE PLANE CONDUCT
823 669 208 759 350

Request 74 (D.M. Yates) (7)
PREDICTING THE PATHS OF ELECTRONS MOVING IN A VARYING MAGNETIC FIELD
PREDICT PATH ELECTRON MOVING VARY MAGNETIC FIELD
431 906 87 571 949 166 703

Request 75 (J.R. Parks) (6)
ACTIVE CONSTANT VOLTAGE TRANSFORMER FOR SIGNAL DISTRIBUTION
ACTIVE CONSTAN VOLT TRANSFOR SIGNAL DISTRIBU
487 352 959 312 613 80

Request 76 (J.R. Parks) (4)
ACTIVE AUDIO FREQUENCY FILTER WITH VARIABLE CUT OFF SLOPE
ACTIVE FREQUENC FILT VARIABLE
487 108 867 323

Request 77 (J.R. Parks) (4)
VARIABLE ULTRA HIGH FREQUENCY ATTENUATORS
VARIABLE HIGH FREQUENC ATTENUAT
323 879 108 23

Request 78 (J.R. Parks) (5)
FAULTY OPERATION OF FEEDBACK TIME BASES
FAULT OPERATIO FEEDBACK TIME BASE
[illegible]1 201 101 942 852

Request 79 (P.J. Pobgee) (3)
VARIABLE CAPACITANCE AMPLIFIERS
VARIABLE CAPACIT AMPLIF
323 341 491

Request 80 (P.J. Pobgee) (3)
TRANSISTOR SWEEP GENERATORS
TRANSIST SWEEP GENERAT
314 806 388

Request 81 (P.J. Pobgee) (3)
ADVANTAGES OF PARAMETRIC AMPLIFIERS
ADVANTAG PARAMET AMPLIF
5 419 491

Request 82 (P.J. Pobgee) (3)
OPTIMISING LINEAR NETWORKS
OPTIM LINEAR NETWORK
754 559 412

Request 83 (P.J. Pobgee) (4)
TRANSISTOR PHASE SPLITTING CIRCUITS
TRANSIST PHAS SPLITT CIRCUIT
314 908 618 345

Request 84 (E.A. Newman) (5)
PLEASE SUPPLY INFORMATION PERTINENT TO THE USE OF SURFACE PRETREATMENT TO PREVENT SECONDARY EMMISSION EFFECTS
INFORMAT SURFACE SECONDAR EMISSI
136 469 262 531

IN VALVES
VALVE
814

Request 85 (E.A. Newman) (7)
I WISH TO HAVE DATA ABOUT THE DESIGN OF MECHANICAL BAND PASS FILTERS FOR GOOD PASS CHARACTERISTICS
DATA DESIGN MECHANIC BAND FILT PASS CHARACTE
848 520 173 831 867 905 39

Request 86 (E.A. Newman) (6)
MECHANISMS WHEREBY TRANSMISSION AT HIGH FREQUENCIES IS AFFECTED BY WEATHER AND TIME OF DAY
MECHANIS TRANSMI HIGH FREQUENC TIME DAY
174 476 879 108 942 969

Request 87 (E.A. Newman) (2)
COULD YOU PLEASE GIVE ME ARTICLES ABOUT THE POSSIBILITIES OF GETTING RECTIFICATION USING METALLIC DEVICES
RECTIF METAL
599 743

Request 88 (E.A. Newman) (7)
I WISH TO CALCULATE THE INDUCTANCE AND LOSS IN COILS MADE USING PRINTED CIRCUIT OR OTHER MINIATURIZATION
CALCULAT INDUCTAN LOSS COIL PRINT CIRCUIT MINIATUR
35 134 893 841 764 345 181

IDEAS. A SUITABLE ARTICLE PLEASE

Request 89 (R.S. Watson) (5)
ANALYSIS OF NONLINEAR SYSTEMS USING PHASE PLANE TECHNIQUES
ANALY NONLINEA SYSTEM PHAS PLANE
658 192 632 908 759

Request 90 (R.S. Watson) (7)
MINIATURE LOW NOISE HIGH GAIN HIGH IMPEDANCE AMPLIFIERS
MINIATUR LOW NOISE GAIN HIGH IMPEDANC AMPLIF
181 982 900 874 879 128 491

Request 91 (R.S. Watson) (5)
CONTROL CHARACTERISTICS OF SAMPLING SERVO SYSTEMS
CONTROL CHARACTE SAMPL SERVO SYSTEM
356 39 779 782 632

Request 92 (R.S. Watson) (5)
POWER SPECTRAL DENSITY DISTRIBUTION OBTAINED USING ANALOGUE TECHNIQUES
POWER SPECTR DENSIT DISTRIBU ANALOGUE
763 616 519 80 10

Request 93 (R.S. Watson) (7)
HIGH STABILITY HIGH INPUT IMPEDANCE TRANSISTORISED ANALOGUE AMPLIFIER
STABILI HIGH INPUT IMPEDANC TRANSIST ANALOGUE AMPLIF
461 879 724 128 314 10 491

Request 94 (P.R. Stuart) (6)
ELECTRONIC SPECIFIC HEAT OF A SUPERCONDUCTOR SHOWING A DISCONTINUITY AT THE SUPERCONDUCTING CRITICAL
ELECTRON HEAT SUPERCON DISCONTI (SUPERCON) CRITICAL
87 878 291 77 (291) 64

TEMPERATURE
TEMPERAT
296

Request 95 (P.R. Stuart) (8)
AN ABSTRACT ON THE FIELD DISTRIBUTION SURROUNDING A CHARGED THIN CIRCULAR DISC RESTING ON AN INFINITE
FIELD DISTRIBU CHARG THIN CIRCULAR DIS INFINIT
703 80 669 941 40 970 395

DIELECTRIC SLAB
DIELECTR
70

Request 96 (P.R. Stuart) (4)
TUNNEL DIODE CONSTRUCTION AND ITS ELECTRICAL CHARACTERISTICS EXPLAINED
TUNNEL DIODE ELECTRIC CHARACTE
642 689 84 39

Request 97 (P.R. Stuart) (5)
ELECTRONIC DENSITY OF STATES AT THE SURFACE OF A SEMICONDUCTOR COMPARED WITH THAT AT DEPTH
ELECTRON DENSIT SURFACE SEMICOND DEPTH
87 519 469 264 685

Request 98 (P.R. Stuart) (6)
RESISTIVITY OF METALLIC THIN FILMS RELATED TO SURFACE ROUGHNESS
RESISTIV METAL THIN FILM SURFACE ROUGHNES
251 743 941 866 469 259

Request 99 (D.M. Yates) (9)
THE PHENOMENON OF RADIATION CAUSED BY CHARGED PARTICLES MOVING IN VARYING ELECTRIC AND MAGNETIC FIELDS
PHENOMEN RADIATIO CHARG PARTICLE MOVING VARY ELECTRIC MAGNETIC FIELD
213 235 669 208 571 949 84 166 703

APPENDIX B2

Numbers and distribution of key-word in requests 0-99

(   ) = cancelled request

Request Number	0	1	2	3	4	5	6	7	8	9
0	6	(10)	5	6	9	5	(3)	5	(4)	4
10	7	7	7	5	3	7	8	6	7	7
20	9	6	6	6	7	6	8	4	7	5
30	4	(5)	4	3	5	9	5	10	7	4
40	5	5	9	3	6	(5)	(4)	3	4	6
50	3	7	6	6	5	6	3	9	7	6
60	6	5	4	6	6	7	(8)	8	5	3
70	4	7	5	5	7	6	4	4	5	3
80	3	3	3	4	5	7	6	2	7	5
90	7	5	5	7	6	8	4	5	6	9

Distribution

Number of Keywords

	1	2	3	4	5	6	7	8	9	10	
Requests		87	(6)	(8)	2	0	10	16	4	(1)	
			14	9	5	3	11	26	20	37	
			33	27	7	17	12	(66)	35		
			43	30	13	21	15	67	42		
			47	32	29	22	18	95	57		
			50	39	(31)	23	19		99		
			56	(46)	34	25	24				
			69	48	36	44	28				
			79	62	40	49	38				
			80	70	41	52	51				
			81	76	(45)	53	58				
			82	77	54	55	65				
				83	61	59	71				
				96	68	60	74				
					72	63	85				
					73	64	88				
					78	75	90				
					84	86	93				
					89	94					
					91	98					
					92						
					97						
Number of requests	0	1	11(+1)	12(+2)	20(+2)	20	18	4(+1)	6	1(+1)	= 93

93 requests,
Mean 5.75 key-words
Mode 5.5 key-words

Standard Deviation
1.6 key-words

APPENDIX B3

Subsets of Requests

(i) Cancelled Requests

7 of the original 100 requests were cancelled at various stages of the work.

Request 1. 'Super-regenerative' is a single word in the dictionary and the abstracts, but hyphens were coded as spaces throughout our work.

Request 6. "Testing" is not in the dictionary.

Request 8. Request garbled in typing.

Request 31. 'Ozone' is not in the dictionary.

Request 45. Requestor not available for assessments.

Request 46. Dictionary mistakenly does not accept plural of 'disc'.

Request 66. 'Non-linear', as with Request 1.

(ii) The '55' set (report VI.2.6)

The following 55 requests had one or more words underlined:-

0, 2, 4, 7, 10-12, 14-19, 22-24, 26-30, 32, 33, 35, 36, 42-44, 47-50, 58, 59, 61, 62, 64, 65, 67, 68, 71, 74-79, 81, 89, 94-99.

(iii) The '34' set (report, VIII.2)

The following 34 requests retrieved only 0-4 relevant documents in Run 13 at K' ≃ 50:-

3, 5, 7, 11-14, 35, 38, 40, 43, 49-51, 54-56, 60, 64, 65, 67, 71-74, 77, 78, 81, 87, 88, 92, 94, 96, 99.

(iv) The '17' set (report, VIII.1)

The following 17 requests were reckoned less well formulated:-

5, 12, 13, 34, 43, 52-55, 60, 62, 76, 84-86, 94, 96.

(v) Four Request Sets of Increasing Generality (report VIII.4)

Gen 1 (0-8 known relevant documents) (26 requests)
5, 7, 11, 12, 14, 32, 34, 37, 38, 40, 43, 49, 50, 51, 54, 55, 56
65, 67, 72, 73, 77, 83, 88, 92, 94

Gen 2 (9-17 known relevant documents) (21 requests)
3, 9, 13, 21, 22, 33, 35, 36, 42, 57, 59, 60, 61, 63, 64, 76, 87,
91, 95, 96, 99

Gen 3 (18-29 known relevant documents) (23 requests)
2, 18, 19, 20, 23, 26, 30, 41, 44, 58, 69, 70, 71, 74, 78, 80, 84,
85, 86, 89, 90, 93, 97

Gen 4 (30-84 known relevant documents) (23 requests)
0, 4, 10, 15, 16, 17, 24, 25, 27, 28, 29, 39, 47, 48, 52, 53, 62, 68
75, 79, 81, 82, 98.

(vi) <u>The 'Subject Index' set</u> (report, VII.4)

Subject Index searches were carried out for requests

0, 18, 22, 38, 41, 57, 58, 70, 74, 75, 83, 94.

(vii) <u>The 'Virtually Exhaustive Search' set</u> (report, VII.4)

This involved requests

10, 38, 91.

APPENDIX B4

Summary Description of Strategies (report, VI.1).

Mnemonic	Run	Formula for Coordination Value
Basic Key Word Stem Strategy		
KWS	13	Number of Key-word Stems common to abstract and request
KWS with Automatic Weighting		
AWKWS	22	Weight = 2 if word not present in 80% of output for Run 0 (i.e. Run 13 at K = 50, K' ≈ 35)
Descriptor Strategies		
		Number of Descriptors in common, using:-
ARM	16	A R Meetham's non-overlapping clusters from G3, (C2)
MCS01	14	P Vaswani's 1178 overlapping clusters, (C3),
MCS11	20	P Vaswani's 725 overlapping clusters, (C5),
RJR	19	R J Reason's 702 slightly overlapping clusters from G3, (C4)
Descriptors with Representation		
ARMSR	17	2^4. Run 16 Representation + Run 16 Coordination
SR14	6	2^7. Run 14 Representation + Run 14 Coordination
PDR14	11	2^7. Run 14 Proportional Representation + Run 14 Coordination, where a request word is represented if a third of its descriptors are assigned to the abstract
Association Strategies		
EAG3	15	Number of Request Words associated with some abstract word via G3
EAG4	18	As 15 but with G4
EARG4	23	Number of Associations via G4
Combinations of Two Strategies		
13T14	9	Run 13 Coordination times Run 14 Coordination
13W14	28	2^7. Run 13 Coordination + Run 14 Coordination, effectively subdividing the strata of Run 13
Strategies with Underlining		
U14	21	Run 14 with underlining
U11	24	Run 11 with underlining
U13	25	Run 13 with underlining
Searches Largely Manual		
SI	26	Subject Indexes - No Coordination
EXHSEA	27	Virtually Exhaustive Search - No Coordination

Discarded

ARM- 1 As Run 16 but ignoring isolated words as descriptors.

Note 1. Runs 0 2 3 4 5 6 7 8 10 12 were
Runs 13 15 16 17 14 19 20 21 22 respectively at $K = 50$, rather than $K' \simeq 50$.

Note 2. Run 24 is fully asse ed up to and including $K = 18$, $K' \simeq 15$
25 $K = 27$, $K' \simeq 17$
28 $K = 42$, $K' \simeq 40$

but careful use may be made of these runs beyond this point. For example, Run 28 and Run 13 yield identical total numbers, or nearly, of relevant documents at $K' \simeq 50$ as well as at $K' = 20, 30, 40$.

APPENDIX B5

Strategy Features

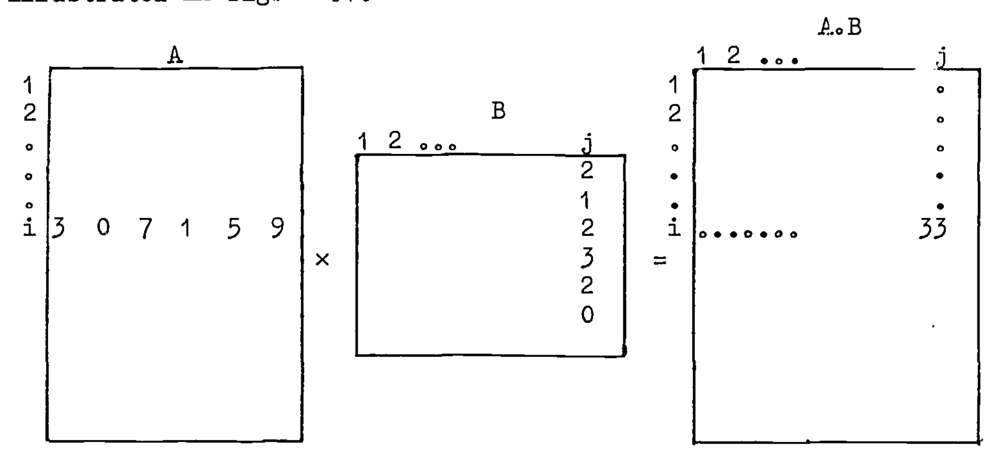
Run	Mnemonic	Maximum Coordination attained	Values of K yielding as near as possible the given K'								
			$K' \simeq 20$	$K' \simeq 30$	$K' \simeq 40$	$K' \simeq 45$	$K' \simeq 46$	$K' \simeq 47$	$K' \simeq 48$	$K' \simeq 49$	$K' \simeq 50$
6	SR14	1220	20	30	41	45	46	47	48	49	49
9	13T14	552	22	31	42	45	47	48	48	48	49
11	PDR14	1104	20	30	41	45	46	47	48	49	50
13	KWS	8	28	43	58	67	68	69	69	70	71
14	MCS01	99	22	35	43	49	50	52	53	54	55
15	EAG3	8	27	42	53	60	61	64	69	70	73
16	ARM	8	27	44	67	75	75	75	76	77	79
17	ARMSR	136	26	44	62	72	73	75	75	76	76
18	EAG4		30	43	53	61	62	63	63	64	65
19	RJR	10	29	44	55	59	60	61	64	65	66
20	MCS11	11	25	46	56	64	67	70	70	73	73
21	U14	99	25	37	46	53	54	55	56	59	62
22	AWKWS	13	28	39	49	59	60	61	61	62	62
23	EARG4		21	33	42	49	50	50	51	52	53
24	U11		24	35	44						
25	U13		29	47	66	70	72	73	74	75	81
28	13W14		22	32	42	47	48	49	50	51	52

APPENDIX B6

Computer Implementation of Indexing and Retrieval Operations

Matrix manipulation of data

A matrix is a rectangular or square array of numbers or data elements. We have dealt with them throughout this report, but we have not mentioned operations upon matrices. One used extensively in the organization of computer programs for indexing and retrieval is that of multiplying two matrices. The product, A.B, of two matrices is defined only when the number of columns of matrix A and the number of rows of matrix B are equal. The product is then another matrix having as many rows as A and as many columns as B. The element in row-i and column-j of the product is obtained as the sum of products of corresponding elements in row-i of A and column-j of B. This is illustrated in fig. B6.1.



$3\times2+0\times1+7\times2+1\times3+5\times1\,0+9\times0=33$

Fig. B6.1 Illustration of matrix multiplication

Consider this process in the case where A is a binary abstract/word matrix (i.e. rows corresponding to abstracts, columns to words) and B is a binary word/request matrix. Row-i of A is then a string of 0's and 1's, the 1's representing the particular words in abstract-i, and column-j of B is a similar string of 0's and 1's in which the 1's represent the particular words occurring in request-j. The value of the element in row-i and column-j of the product will then be the number of words in common, i.e. the key word coordination level, for the abstract and request in question. The product matrix contains the key word coordination levels for all abstracts with respect to all requests. The rows of this matrix correspond to abstracts and the columns to requests. The organization of our programs in the computer is such that the matrix of coordination levels produced in

each run is required to have rows corresponding to requests. Having produced the matrix for run 13 as above it is therefore necessary to transpose it so that its columns become rows and vice-versa. Before transferring the lists of retrieved abstracts to the harvest file (Section V.8) it is also necessary to apply the necessary coordination cut-off level for each request (Section VI.1), rejecting abstracts at lower levels.

Three matrix products are formed in the case of a simple descriptor run such as run 14. In the first the abstract/word matrix is multiplied by a word descriptor one to obtain an abstract/descriptor matrix. In accordance with the indexing strategy this is converted to binary form by setting a threshold value of one. We now have a matrix indicating the descriptors by which each abstract is indexed. The descriptor/word matrix (the transpose of the word/descriptor one) is then multiplied by the word/request matrix, again thresholding at one, to give a descriptor/request matrix. Finally the product of the abstract/descriptor and descriptor/request matrices yields an abstract/request one which, as before, requires transposing and application of the cut-off process.

Run 14, an associative run using word connection matrix G3 directly, involves only two products. First the abstract/word matrix is multiplied by G3. This yields a new abstract/word matrix which is converted to binary by thresholding its values at one (for strategy see Section VI.1). This is then multiplied by the word/request matrix to obtain the abstract/request matrix.

The main operations used are thus multiplication, transposition and thresholding.

Data formats

Two formats were adopted for matrices. With the first, used for storage on magnetic tape or for holding individual matrix rows in the core store, each row is stored as a block of successive machine words, the first of which specifies the number of words to follow. The remaining words in a block specify the non-zero elements in the matrix row, each word containing the element value and column number of one element.

Binary matrices with not more than 1,000 rows or columns could be held entirely in the core store by adopting a more economical format. Column numbers of all the non-zero elements of a matrix are held, three per word, in a block of words and are accompanied by a key indicating the starting point within the block of each row.

Matrix multiplication

With one or two exceptions all the multiplication involves binary matrices. For simplicity and processing speed the programs took advantage of this fact. Referring to fig. B6.1 we see that the non-zero elements of row-i of matrix A indicate those rows of matrix B that must be accessed to obtain element-ij of the product. Furthermore computation of every element of row-i of the product requires access to these same rows of matrix B. It is therefore economical to compute all the elements in a row of the product together. For binary matrices it is necessary simply to add together corresponding elements of all the selected rows of matrix B in order to obtain a row of the product matrix.

Two multiplication routines are used working on this principle. One is used when both matrices in a product are small and are held entirely in the core store. The other requires the post-multiplier (the R.H. matrix) to be in the core store and reads successive rows of the pre-multiplier from magnetic tape. Both write successive rows of the product, which is generally non-binary, to magnetic tape.

Matrix thresholding

One program reads a complete non-binary matrix from magnetic tape and stores it in compact binary form in the core store after setting a threshold value one (the commonest value).

Another program accepts any stated value in thresholding a single row of a non-binary matrix held in the core store and leaves the result in the core.

Coordination value cut-off

The abstract/request matrices have 11,571 rows and 100 columns. Transposing a matrix of this size on a computer is a difficult and a slow job, the time increasing with the number of non-zero matrix elements. For this reason the coordination cut-off level for each request is computed and applied before transposing the matrix.

Two programs are used here. We wish to obtain an average output as close as possible to 50 abstracts per request by retrieving as near as possible to some fixed number, K, of abstracts for each (see Section VI.1). The first program forms 100 distributions, one for each request, of the coordination values (i.e. the element values of the abstract /request matrix) and from them computes the value K. The second program makes two passes of the tape containing the abstract/request matrix. During the first the distributions are re-formed and the coordination cut-off level giving an

output closest to K is recorded for each request. On the second tape pass the matrix is thresholded using these separate cut-off values for the different columns and the resulting matrix is written on to magnetic tape.

Matrix transpose

Difficulty only arises in the case of very large matrices such as the abstract/request and abstract/word ones. We will therefore only describe the method used for these. The number of passes of the tape containing the matrix to be transposed depends upon the number of non-zero elements. The first pass is used to find the number of elements in each column. Knowing this and the amount of core store made available for holding the transpose, the program decides how many rows of the transpose to form on the next and subsequent pa s of the tape. The requisite number of passes of the input tape is then made, each time forming a subset of the rows of the transpose in the available core store and then writing them on another magnetic tape.

Retrieval print-out

The texts of all 11,571 abstracts are read from magnetic tape and dumped on to a magnetic disc store giving reasonable random access. Having ied the contents of the request/abstract matrix for the run in hand on to the harvest file, the latter is now accessed to obtain the list of abstract numbers retrieved for each request. The text of each request, held in the core store, is assembled with those of the retrieved abstracts as they are called down from the disc. The complete output for each request is sent in the required format to a magnetic tape. When the entire output for a run is written on tape another program is run which reads it from the tape and sends it to the line-printer. The programs are organized in this way so that the main program, the run time of which is drawn out by the vast number of disc accesses, does not monopolize the line-printer, inhibiting its use by other programs that may be running in the machine in parallel.

Computer time required for indexing and retrieval

It is assumed that the abstract/word matrix the word/request matrix and any required descriptor/word matrix are available in machine readable form.

When a program being run in time sharing mode on KDF9 terminates, three times are printed out:

1. the time for which the program actually occupied the central processor unit,

2. the total elapsed time during which the program was running in parallel with others, and

3. the notional elapsed time, which is an estimate of the time for which the program would run on a machine without time sharing.

In calculating the following times (3) was used.

We consider run 13, the basic key-word strategy, and runs 11 and 23, the best single runs with which to supplement it. Run 11 uses a cluster strategy with proportional representation and run 23 involves expanding requests and abstracts with G4, but does not use clusters (see Section VI.1).

Below we give the total times to run all necessary programs to produce, from the basic matrices mentioned above, a request/abstract matrix (for 93 requests):

Run	Number of matrix operations	total notional elapsed time
13	3	16 mins
11	10	164 "
23	5	69 "

Although the machine times required, having produced a request/abstract matrix, to obtain the final printed output of retrieved material are substantial, they would not be very meaningful figures to quote. This is because the economics of this end of the process depend upon factors such as whether or not the computer is time-shared, what sort of backing store the document texts are held on and whether retrieved texts are printed on-line to the computer or off-line. These times are also independent of retrieval strategy.

In our case the texts of the abstracts are held on magnetic tape and are only dumped on to a magnetic disc when the appropriate program is running. It takes approximately 17 minutes to transfer them to the disc, but as our machine is time-shared that is quite reasonable. In an operational environment the texts might reside permanently on a disc file. The line-printer on our KDF9 can only be used in an on-line mode. Again, having time-sharing, it is not too unreasonable to print extensive quantities of output in this way, although it does monopolize the printer. In an operational system one would try to arrange that large amounts of printing are produced in an off-line mode.

Evaluation Files and Programs

The Harvest File is the central information bank and is arranged thus on magnetic tape:-

There is one block of information on the tape per request.

The block's machine words are each arranged thus,

Flag/Strategy Number/Relevance/Coordination/Abstract,

ordered by strategy and then abstract.

The blocks are overwritable for later insertion of relevances, or other corrections. To lengthen the blocks with the results of a new strategy the file is copied from one of two tapes to the other in turn.

The flag marks abstracts retrieved by no previous strategy.

Flagged abstracts only are sent to request or to avoid duplicate assessment and reduce unnecessary output. The flags are also of use in counting, for example, the numbers of distinct relevant abstracts, or all distinct abstracts.

The harvest file can be reduced to study subsets of requests or smaller cut-offs.

Programs to Run and Evaluate a Typical Strategy

Each step after the first represents a single program. They rely on 4-25K store and on a single pass of each with each input or output tape, and take 3-15 minutes on KDF9.

1. Form coordination for each abstract-request, for example by matrix multiplication or merging two tapes, in order of abstracts.
2. Print out numbers of abstracts in each stratum.
3. List K' for K = 1, 2, 3 ......
4. Cut off 1) at largest K or K' likely to be needed and transpose, i.e. rewrite 1) in order of requests.
5. Suppress permanently cancelled requests as necessary.
6. Add new strategy to Harvest file, mark abstracts not previously retrieved with flag.
7. Print out new abstracts and send to requestors for assessment.
8. Insert assessments of flagged abstracts via paper tape.
9. On Harvest Tape, copy assessments of all unflagged items where known from an earlier run.
10. Replace existing coordinations by 100 - K, where K = minimum K-value needed to retrieve stratum in which abstract occurs. This preserves the strata while enabling any smaller cut-off to be applied by a simple program step.

11. Form one new Harvest File for each desired K', using appropriate K from step 3.

12. Reset Harvest File Flags if earlier runs have had items removed in step 11.

13. Form other Harvest Files with selected requests only.

14. Several Print-Outs, each displaying:

<u>For Fixed Cut Off, Each Run, Each Request</u>

The numbers of relevant, irrelevant, the % precision and recall, also totals and average numbers of relevant per request; the same after suppressing abstracts also found in a stated run (e.g. Run 13) or runs.

15. Several Print-Outs, each displaying:

<u>For K = 1, 2, 3</u> ..... and <u>One Run, One Request Set</u>

K, K', $\frac{K'}{K}$, totals relevant, irrelevant, unassessed, output, Overall Precision and Recall (report VII.5), Average Precision and Recall with Standard Deviations.

<u>Programs to display behaviour of strategy for large cut-offs</u>

16. Collect abstracts known to be relevant from all available sources.

17. Insert coordinations for given strategy from 1).

18. For each stratum display numbers of relevant from 16), and of the remainder, assumed irrelevant, using 2), precision and recall.

19. As 15) as far as Overall Precision and Recall, for K = 1 to 500 (say).

<u>Other Programs</u>

20. For a given Run display request words represented by each abstract.

21. Display words associated with request words via some descriptor.

APPENDIX B7

Performance of Successive Strata in Run 13(KWS). Table RCC00.

Unassessed documents, that is, those not retrieved by standard strategies or by the two manual searches, have been set irrelevant (c.f. VII.4).

Coord-ination	Cumul-ative Output	Rele-vant	Irrel-evant	%Prec-ision
			REQUEST	0
5	14	7	7	50.00
4	101	30	57	34.48
3	564	8	455	1.73
2	1952	1	1387	.07
1	5591	0	3639	.00
0	11571	0	5980	.00
	TOTAL	46	11525	.40
			REQUEST	1
0	11571	0	11571	.00
	TOTAL	0	11571	.00
			REQUEST	2
4	2	1	1	50.00
3	28	8	18	30.77
2	231	9	194	4.43
1	2034	1	1802	.06
0	11571	0	9537	.00
	TOTAL	19	11552	.16
			REQUEST	3
3	14	4	10	28.57
2	242	7	221	3.07
1	2668	4	2422	.16
0	11571	0	8903	.00
	TOTAL	15	11556	.13
			REQUEST	4
7	1	1	0	100.00
6	9	2	6	25.00
5	32	10	13	43.48
4	105	5	68	6.85
3	300	10	185	5.13
2	873	4	569	.70
1	3092	1	2218	.05
0	11571	0	8479	.00
	TOTAL	33	11538	.29
			REQUEST	5
3	7	2	5	28.57
2	125	2	116	1.69
1	1568	1	1442	.07
0	11571	0	10003	.00
	TOTAL	5	11566	.04
			REQUEST	6
0	11571	0	11571	.00
	TOTAL	0	11571	.00
			REQUEST	7
3	12	0	12	.00
2	375	3	360	.83
1	3701	1	3325	.03
0	11571	0	7870	.00
	TOTAL	4	11567	.03
			REQUEST	8
0	11571	0	11571	.00
	TOTAL	0	11571	.00
			REQUEST	9
3	4	3	1	75.00
2	53	3	46	6.12
1	739	4	682	.58
0	11571	0	10832	.00
	TOTAL	10	11561	.09
			REQUEST	10
5	5	2	3	40.00
4	29	9	15	37.50
3	114	44	41	51.76
2	399	17	268	5.96
1	2196	3	1794	.17
0	11571	0	9375	.00
	TOTAL	75	11496	.65
			REQUEST	11
5	1	0	1	.00
4	9	1	7	12.50
3	70	0	61	.00
2	546	0	476	.00
1	3073	0	2527	.00
0	11571	0	8498	.00
	TOTAL	1	11570	.01
			REQUEST	12
5	3	1	2	33.33
4	22	0	19	.00
3	79	1	56	1.75
2	341	0	262	.00
1	2416	0	2075	.00
0	11571	0	9155	.00
	TOTAL	2	11569	.02
			REQUEST	13
4	2	0	2	.00
3	19	4	13	23.53
2	360	7	334	2.05
1	2922	0	2562	.00
0	11571	0	8649	.00
	TOTAL	11	11560	.10

REQUEST 14

3	1	0	1	.00
2	51	0	50	.00
1	1884	4	1829	.22
0	11571	0	9687	.00
	TOTAL	4	11567	.03

REQUEST 15

5	4	2	2	50.00
4	24	11	9	55.00
3	232	26	182	12.50
2	1061	1	828	.12
1	3823	0	2762	.00
0	11571	0	7748	.00
	TOTAL	40	11531	.35

REQUEST 16

5	13	7	6	53.85
4	43	19	11	63.33
3	224	23	158	12.71
2	857	8	625	1.26
1	3003	2	2144	.09
0	11571	0	8568	.00
	TOTAL	59	11512	.51

REQUEST 17

4	3	1	2	33.33
3	52	29	20	59.18
2	324	19	253	6.99
1	2452	6	2122	.28
0	11571	1	9118	.01
	TOTAL	56	11515	.48

REQUEST 18

5	2	0	2	.00
4	13	7	4	63.64
3	107	10	84	10.64
2	611	9	495	1.79
1	2637	6	2020	.30
0	11571	0	8934	.00
	TOTAL	32	11539	.28

REQUEST 19

4	19	16	3	84.21
3	167	10	138	6.76
2	964	2	795	.25
1	4059	0	3095	.00
0	11571	0	7512	.00
	TOTAL	28	11543	.24

REQUEST 20

7	1	0	1	.00
6	9	2	6	25.00
5	27	5	13	27.78
4	118	7	84	7.69
3	395	7	270	2.53
2	1414	2	1017	.20
1	4097	0	2683	.00
0	11571	0	7474	.00
	TOTAL	23	11548	.20

REQUEST 21

5	5	0	5	.00
4	43	6	32	15.79
3	275	5	227	2.16
2	1163	2	886	.23
1	3296	0	2133	.00
0	11571	0	8275	.00
	TOTAL	13	11558	.11

REQUEST 22

5	2	2	0	100.00
4	18	7	9	43.75
3	243	9	216	4.00
2	1127	4	880	.45
1	3655	1	2527	.04
0	11571	1	7915	.01
	TOTAL	24	11547	.21

REQUEST 23

4	3	3	0	100.00
3	30	8	19	29.63
2	366	12	324	3.57
1	1991	0	1625	.00
0	11571	0	9580	.00
	TOTAL	23	11548	.20

REQUEST 24

6	2	2	0	100.00
5	15	8	5	61.54
4	70	21	34	38.18
3	237	16	151	9.58
2	792	3	552	.54
1	2759	0	1967	.00
0	11571	0	8812	.00
	TOTAL	50	11521	.43

REQUEST 25

5	3	2	1	66.67
4	18	10	5	66.67
3	91	31	42	42.47
2	661	19	551	3.33
1	3354	3	2690	.11
0	11571	0	8217	.00
	TOTAL	65	11506	.56

REQUEST 26

7	1	1	0	100.00
6	2	0	1	.00
5	6	2	2	50.00
4	24	4	14	22.22
3	85	5	56	8.20
2	440	7	348	1.97
1	2711	1	2270	.04
0	11571	0	8860	.0
	TOTAL	20	11551	.1

REQUEST 27

4	3	3	0	100.0
3	55	15	37	28.8
2	374	14	305	4.3
1	1621	7	1240	.5
0	11571	0	9950	.0
	TOTAL	39	11532	.3

REQUEST 28

4	17	4	13	23.53
3	122	12	93	11.43
2	599	55	422	11.53
1	2491	2	1890	.11
0	11571	0	9080	.00
	TOTAL	73	11498	.63

REQUEST 29

4	2	2	0	100.00
3	34	8	24	25.00
2	291	43	214	16.73
1	1867	0	1576	.00
0	11571	0	9704	.00
	TOTAL	53	11518	.46

REQUEST 30

3	7	5	2	71.43
2	70	12	51	19.05
1	1014	11	933	1.17
0	11571	0	10557	.00
	TOTAL	28	11543	.24

REQUEST 31

0	11571	0	11571	.00
	TOTAL	0	11571	.00

REQUEST 32

3	4	1	3	25.00
2	73	5	64	7.25
1	765	2	690	.29
0	11571	0	10806	.00
	TOTAL	8	11563	.07

REQUEST 33

3	3	3	0	100.00
2	83	8	72	10.00
1	1138	1	1054	.09
0	11571	0	10433	.00
	TOTAL	12	11559	.10

REQUEST 34

5	2	1	1	50.00
4	13	1	10	9.09
3	102	3	86	3.37
2	522	2	418	.48
1	2950	0	2428	.00
0	11571	0	8621	.00
	TOTAL	7	11564	.06

REQUEST 35

6	1	1	0	100.00
5	5	2	2	50.00
4	47	0	42	.00
3	341	3	291	1.02
2	1641	4	1296	.31
1	5477	1	3835	.03
0	11571	0	6094	.00
	TOTAL	11	11560	.10

REQUEST 36

4	3	3	0	100.00
3	15	4	8	33.33
2	105	6	84	6.67
1	1321	0	1216	.00
0	11571	0	10250	.00
	TOTAL	13	11558	.11

REQUEST 37

6	5	1	4	20.00
5	33	2	26	7.14
4	104	2	69	2.82
3	391	3	284	1.05
2	1640	0	1249	.00
1	5046	0	3406	.00
0	11571	0	6525	.00
	TOTAL	8	11563	.07

REQUEST 38

6	1	0	1	.00
5	12	2	9	18.18
4	70	2	56	3.45
3	264	1	193	.52
2	1008	1	743	.13
1	4384	3	3373	.09
0	11571	0	7187	.00
	TOTAL	9	11562	.08

REQUEST 39

4	2	2	0	100.00
3	44	14	28	33.33
2	339	14	281	4.75
1	2218	0	1879	.00
0	11571	0	9353	.00
	TOTAL	30	11541	.26

REQUEST 40

3	24	1	23	4.17
2	371	6	341	1.73
1	2817	1	2445	.04
0	11571	0	8754	.00
	TOTAL	8	11563	.07

REQUEST 41

5	1	1	0	100.00
4	9	6	2	75.00
3	27	6	12	33.33
2	168	8	133	5.67
1	1504	4	1332	.30
0	11571	0	10067	.00
	TOTAL	25	11546	.22

REQUEST 42

7	1	1	0	100.00
6	2	0	1	.00
5	7	2	3	40.00
4	15	2	6	25.0[illegible]
3	56	6	35	14.6[illegible]
2	382	1	325	.31
1	2329	0	1947	.00
0	11571	0	9242	.00
	TOTAL	12	11559	.10

REQUEST 43

2	32	2	30	6.25
1	850	3	815	.37
0	11571	0	10721	.00
	TOTAL	5	11566	.04

REQUEST 44

4	12	8	4	66.67
3	33	7	14	33.33
2	197	12	152	7.32
1	1047	2	848	.24
0	11571	0	10524	.00
	TOTAL	29	11542	.25

REQUEST 45

0	11571	0	11571	.00
	TOTAL	0	11571	.00

REQUEST 46

0	11571	0	11571	.00
	TOTAL	0	11571	.00

REQUEST 47

3	2	1	1	50.00
2	86	41	43	48.81
1	796	32	678	4.51
0	11571	10	10765	.09
	TOTAL	84	11487	.73

REQUEST 48

4	10	10	0	100.00
3	65	16	39	29.09
2	282	7	210	3.23
1	1751	3	1466	.20
0	11571	0	9820	.00
	TOTAL	36	11535	.31

REQUEST 49

4	3	3	0	100.00
3	34	1	30	3.23
2	317	3	280	1.06
1	1134	0	817	.00
0	11571	0	10437	.00
	TOTAL	7	11564	.06

REQUEST 50

3	4	3	1	75.00
2	165	3	158	1.86
1	772	2	605	.33
0	11571	0	10799	.00
	TOTAL	8	11563	.07

REQUEST 51

3	9	1	8	11.11
2	163	2	152	1.30
1	1683	2	1518	.13
0	11571	0	9888	.00
	TOTAL	5	11566	.04

REQUEST 52

5	1	0	1	.00
4	8	5	2	71.43
3	37	9	20	31.03
2	240	27	176	13.30
1	1441	7	1194	.58
0	11571	0	10130	.00
	TOTAL	48	11523	.41

REQUEST 53

4	20	8	12	40.00
3	158	21	117	15.22
2	750	3	589	.51
1	3160	0	2410	.00
0	11571	0	8411	.00
	TOTAL	32	11539	.28

REQUEST 54

4	3	1	2	33.33
3	38	2	33	5.71
[illegible]	375	0	337	.00
[illegible]	2578	0	2203	.00
[illegible]	11571	0	8993	.00
	TOTAL	3	11568	.03

REQUEST 55

4	5	3	2	60.00
3	39	1	33	2.94
2	583	1	543	.18
1	3137	2	2552	.08
0	11571	0	8434	.00
	TOTAL	7	11564	.06

REQUEST 56

3	7	0	7	.00
2	214	1	206	.48
1	1848	0	1634	.00
0	11571	0	9723	.00
	TOTAL	1	11570	.01

REQUEST 57

5	4	2	2	50.00
4	35	3	28	9.68
3	195	7	153	4.38
2	1135	4	936	.43
1	4374	0	3239	.00
0	11571	0	7197	.00
	TOTAL	16	11555	.14

REQUEST 58

5	1	0	1	.00
4	13	4	8	33.33
3	73	10	50	16.67
2	572	6	493	1.20
1	4038	5	3461	.14
0	11571	0	7533	.00
	TOTAL	25	11546	.22

REQUEST 59

5	1	0	1	.00
4	6	1	4	20.00
3	117	13	98	11.71
2	1091	3	971	.31
1	4565	0	3474	.00
0	11571	0	7006	.00
	TOTAL	17	11554	.15

REQUEST 60

3	8	1	7	12.50
2	210	8	194	3.96
1	2118	1	1907	.05
0	11571	0	9453	.00
	TOTAL	10	11561	.09

REQUEST 61

4	1	0	1	.00
3	53	10	42	19.23
2	389	2	334	.60
1	2456	1	2066	.05
0	11571	0	9115	.00
	TOTAL	13	11558	.11

REQUEST 62

3	34	22	12	64.71
2	399	12	353	3.29
1	3151	5	2747	.18
0	11571	0	8420	.00
	TOTAL	39	11532	.34

REQUEST 63

4	18	0	18	.00
3	68	6	44	12.00
2	272	4	200	1.96
1	1814	0	1542	.00
0	11571	0	9757	.00
	TOTAL	10	11561	.09

REQUEST 64

4	4	1	3	25.00
3	74	2	68	2.86
2	539	7	458	1.51
1	3388	0	2849	.00
0	11571	0	8183	.00
	TOTAL	10	11561	.09

REQUEST 65

4	6	0	6	.00
3	50	0	44	.00
2	502	1	451	.22
1	3009	0	2507	.00
0	11571	0	8562	.00
	TOTAL	1	11570	.01

REQUEST 66

0	11571	0	11571	.00
	TOTAL	0	11571	.00

REQUEST 67

6	3	0	3	.00
5	14	0	11	.00
4	80	1	65	1.52
3	415	2	333	.60
2	1633	0	1218	.00
1	4696	0	3053	.00
0	11571	0	6875	.00
	TOTAL	3	11568	.03

REQUEST 68

4	11	5	6	45.45
3	123	26	86	23.21
2	811	13	675	1.89
1	3614	0	2803	.00
0	11571	0	7957	.00
	TOTAL	44	11527	.38

REQUEST 69

3	4	3	1	75.00
2	39	15	20	42.86
1	845	3	803	.37
0	11571	0	10726	.00
	TOTAL	21	11550	.18

REQUEST 70

4	2	2	0	100.00
3	75	17	56	23.29
2	210	6	129	4.44
1	1375	5	1160	.43
0	11571	0	10196	.00
	TOTAL	30	11541	.26

REQUEST 71

4	25	0	25	.00
3	209	17	167	9.24
2	1022	5	808	.62
1	4036	0	3014	.00
0	11571	0	7535	.00
	TOTAL	22	11549	.19

REQUEST 72

4	2	1	1	50.00
3	18	1	15	6.25
2	231	1	212	.47
1	2601	1	2369	.04
0	11571	0	8970	.00
	TOTAL	4	11567	.03

REQUEST 73

4	1	0	1	.00
3	19	4	14	22.22
2	260	4	237	1.66
1	1496	0	1236	.00
0	11571	0	10075	.00
	TOTAL	8	11563	.07

REQUEST 74

5	1	0	1	.00
4	25	4	20	16.67
3	251	12	214	5.31
2	1115	11	853	1.27
1	3663	0	2548	.00
0	11571	0	7908	.00
	TOTAL	27	11544	.23

REQUEST 75

4	1	1	0	100.00
3	23	5	17	22.73
2	297	15	259	5.47
1	2493	12	2184	.55
0	11571	0	9078	.00
	TOTAL	33	11538	.29

REQUEST 76

4	1	1	0	100.00
3	18	7	10	41.18
2	282	5	259	1.89
1	2532	1	2249	.04
0	11571	0	9039	.00
	TOTAL	14	11557	.12

REQUEST 77

4	2	1	1	50.00
3	40	3	35	7.89
2	573	1	532	.19
1	3039	0	2466	.00
0	11571	0	8532	.00
	TOTAL	5	11566	.04

REQUEST 78

3	8	3	5	37.50
2	199	14	177	7.33
1	1874	4	1671	.24
0	11571	0	9697	.00
	TOTAL	21	11550	.18

REQUEST 79

3	18	17	1	94.44
2	165	15	132	10.20
1	1773	0	1608	.00
0	11571	0	9798	.00
	TOTAL	32	11539	.28

REQUEST 80

2	45	13	32	28.89
1	1064	5	1014	.49
0	11571	0	10507	.00
	TOTAL	18	11553	.16

REQUEST 81

3	3	3	0	100.00
2	220	17	200	7.83
1	1584	11	1353	.81
0	11571	1	9986	.01
	TOTAL	32	11539	.28

REQUEST 82

3	1	1	0	100.00
2	94	62	31	66.67
1	1054	3	957	.31
0	11571	0	10517	.00
	TOTAL	66	11505	.57

REQUEST 83

3	19	6	13	31.58
2	484	2	463	.43
1	2417	1	1932	.05
0	11571	0	9154	.00
	TOTAL	9	11562	.08

REQUEST 84

3	24	4	20	16.67
2	115	18	73	19.78
1	1127	2	1010	.20
0	11571	0	10444	.00
	TOTAL	24	11547	.21

REQUEST 85

6	3	0	3	.00
5	22	5	14	26.32
4	81	9	50	15.25
3	221	8	132	5.71
2	607	3	383	.78
1	2691	0	2084	.00
0	11571	0	8880	.00
	TOTAL	25	11546	.22

REQUEST 86

4	8	1	7	12.50
3	103	19	76	20.00
2	808	9	696	1.28
1	3741	0	2933	.00
0	11571	0	7830	.00
	TOTAL	29	11542	.25

REQUEST 87

2	11	4	7	36.36
1	464	8	445	1.77
0	11571	1	11106	.01
	TOTAL	13	11558	.11

REQUEST 88

4	5	0	5	.00
3	31	1	25	3.85
2	333	3	299	.99
1	2786	0	2453	.00
0	11571	0	8785	.00
	TOTAL	4	11567	.03

REQUEST 89

4	3	1	2	33.33
3	57	9	45	16.67
2	393	12	324	3.57
1	2812	0	2419	.00
0	11571	0	8759	.00
	TOTAL	22	11549	.19

REQUEST 90

5	8	1	7	12.50
4	62	12	42	22.22
3	244	10	172	5.49
2	886	2	640	.31
1	3364	1	2477	.04
0	11571	0	8207	.00
	TOTAL	26	11545	.22

			REQUEST	91
4	4	0	4	.00
3	37	9	24	27.27
2	282	9	236	3.67
1	2078	5	1791	.28
0	11571	0	9493	.00
	TOTAL	23	11548	.20

			REQUEST	92
3	12	2	10	16.67
2	231	4	215	1.83
1	2076	2	1843	.11
0	11571	0	9495	.00
	TOTAL	8	11563	.07

			REQUEST	93
6	2	1	1	50.00
5	13	2	9	18.18
4	47	2	32	5.88
3	217	13	157	7.65
2	856	2	637	.31
1	3362	0	2506	.00
0	11571	0	8209	.00
	TOTAL	20	11551	.17

			REQUEST	94
3	13	3	10	23.08
2	212	1	198	.50
1	2206	3	1991	.15
0	11571	0	9365	.00
	TOTAL	7	11564	.06

			REQUEST	95
5	1	1	0	100.00
4	8	3	4	42.86
3	68	2	58	3.33
2	524	3	453	.66
1	2659	1	2134	.05
0	11571	0	8912	.00
	TOTAL	10	11561	.09

			REQUEST	96
3	13	0	13	.00
2	149	12	124	8.82
1	1758	0	1609	.00
0	11571	0	9813	.00
	TOTAL	12	11559	.10

			REQUEST	97
4	1	0	1	.00
3	17	5	11	31.25
2	453	18	418	4.13
1	2107	4	1650	.24
0	.1571	2	9462	.02
	TOTAL	29	11542	.25

			REQUEST	98
4	6	2	4	33.33
3	29	6	17	26.09
2	165	26	110	19.12
1	756	1	590	.17
0	11571	0	10815	.00
	TOTAL	35	11536	.30

			REQUEST	99
6	5	1	4	20.00
5	33	3	25	10.71
4	117	5	79	5.95
3	408	5	286	1.72
2	1312	0	904	.00
1	3561	0	2249	.00
0	11571	0	8010	.00
	TOTAL	14	11557	.12

APPENDIX B8

Performance of Typical Strategies for Varying Cut-off K

See VIII.2,3,4,5 for explanation of the tables.
S.D. = Standard Deviation of the percentages averaged in the immediately preceding columns.

93 REQUESTS, RUN 13(KWS). TABLE RPK(93,13).

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Not Assessed	Output
0	.00	.000	0	0	0	0
1	.22	.220	9	11	0	20
2	.94	.470	49	38	0	87
3	1.60	.533	71	78	0	149
4	2.05	.513	83	108	0	191
5	2.74	.548	102	153	0	255
6	3.43	.572	132	187	0	319
7	4.35	.621	154	251	0	405
8	5.14	.643	179	299	0	478
9	5.58	.620	191	328	0	519
10	6.86	.686	246	392	0	638
11	7.96	.724	280	460	0	740
12	8.04	.670	282	466	0	748
13	9.47	.728	2[illegible]7	584	0	831
14	9.73	.695	301	604	0	905
15	9.95	.663	312	613	0	925
16	10.42	.651	324	645	0	969
17	11.05	.650	334	694	0	1028
18	11.92	.662	371	738	0	1109
19	13.27	.698	392	842	0	1234
20	14.20	.710	400	921	0	1321
21	15.18	.723	421	991	0	1412
22	15.97	.726	439	1046	0	1485
23	17.35	.754	469	1145	0	1614
24	17.81	.742	483	1173	0	1656
25	18.22	.729	489	1205	0	1694
26	18.22	.701	489	1205	0	1694
27	19.20	.711	489	1297	0	1786
28	20.29	.725	528	1359	0	1887
29	21.61	.745	550	1460	0	2010
30	22.17	.739	565	1497	0	2062
31	23.12	.746	576	1574	0	2150
32	23.12	.723	576	1574	0	2150
33	23.12	.701	576	1574	0	2150
34	23.12	.680	576	1574	0	2150
35	23.12	.661	576	1574	0	2150
36	24.[illegible]4	.671	594	1651	0	2245
37	24.14	.652	594	1651	0	2245
38	24.73	.651	610	1690	0	2300
39	27.58	.707	646	1919	0	2565
40	28.99	.725	648	2048	0	2696
41	28.99	.707	648	2048	0	2696
42	29.61	.705	650	2104	0	2754
43	30.20	.702	671	2138	0	2809
44	32.25	.733	695	2304	0	2999

Cut-off K	Average Output K′	Overall %Prec-ision	Overall %Known Recall	Average %Prec-ision	Average %Known Recall	S.D. %Prec-ision	S.D. %Known Recall
0	.00	.00	.00	.00	.00	.00	.00
1	.22	45.00	.43	9.68	.55	29.57	1.90
2	.94	56.32	2.34	25.81	3.93	39.75	8.90
3	1.60	47.65	3.40	31.34	5.94	39.71	10.63
4	2.05	43.46	3.97	32.53	6.88	38.87	11.38
5	2.74	40.00	4.88	33.51	8.15	37.71	11.98
6	3.43	41.38	6.32	34.87	10.40	36.63	15.47
7	4.35	38.02	7.37	36.77	11.79	36.00	15.80
8	5.14	37.45	8.56	37.86	12.91	35.30	15.78
9	5.58	36.80	9.14	37.76	13.14	34.64	15.73
10	6.86	38.56	11.77	39.16	16.10	33.37	18.37
11	7.96	37.84	13.40	39.09	18.04	32.25	18.89
12	8.04	37.70	13.49	38.99	18.22	32.25	19.03
13	9.47	33.71	14.21	38.35	18.91	31.40	18.68
14	9.73	33.26	14.40	38.52	19.07	31.23	18.58
15	9.95	33.73	14.93	38.57	19.37	31.25	18.58
16	10.42	33.44	15.50	38.15	20.04	31.23	18.80
17	11.05	32.49	15.98	37.54	20.84	30.48	18.98
18	11.92	33.45	17.75	38.15	21.76	30.37	19.05
19	13.27	31.77	18.76	36.20	22.84	28.72	19.25
20	14.20	30.28	19.14	35.61	23.54	28.92	19.20
21	15.18	29.82	20.14	35.68	25.00	28.85	20.62
22	15.97	29.56	21.00	34.94	26.41	28.66	22.22
23	17.35	29.06	22.44	34.22	27.81	28.11	22.77
24	17.81	29.17	23.11	33.54	28.31	27.26	22.81
25	18.22	28.87	23.40	33.69	28.80	27.12	22.69
26	18.22	28.87	23.40	33.69	28.60	27.12	22.69
27	19.20	27.38	23.40	33.11	28.80	27.12	22.69
28	20.29	27.98	25.26	33.58	30.19	27.06	22.97
29	21.61	27.36	26.32	32.96	30.86	26.87	23.14
30	22.17	27.40	27.03	32.24	31.27	25.95	23.06
31	23.12	26.79	27.56	31.94	31.82	26.07	22.89
32	23.12	26.79	27.56	31.94	31.82	26.07	22.89
33	23.12	26.79	27.56	31.94	31.82	26.07	22.89
34	23.12	26.79	27.56	31.94	31.82	26.07	22.89
35	23.12	26.79	27.56	31.94	31.82	26.07	22.89
36	24.14	26.46	28.42	31.88	32.85	26.04	23.56
37	24.14	26.46	28.42	31.88	32.85	26.04	23.56
38	24.73	26.52	29.19	31.23	33.33	25.07	23.90
39	27.58	25.19	30.91	29.30	35.29	23.76	24.42
40	28.99	24.04	31.00	28.97	35.50	24.00	24.28
41	28.99	24.04	31.00	28.97	35.50	24.00	24.28
42	29.61	23.60	31.10	28.85	35.74	24.08	24.26
43	30.20	23.89	32.11	28.61	36.19	23.82	24.35
44	32.25	23.17	33.25	27.65	37.98	22.59	24.77

Cut-off K	Average Output K'	K'/K	Rele-vant	Irrel-evant	Not Assessed	Output
45	33.15	.737	736	2347	0	3083
46	33.15	.721	736	2347	0	3083
47	33.15	.705	736	2347	0	3083
48	34.86	.726	799	2443	0	3242
49	34.86	.711	799	2443	0	3242
50	34.86	.697	799	2443	0	3242
51	35.47	.695	800	2499	0	3299
52	36.11	.694	809	2549	0	3358
53	36.11	.681	809	2549	0	3358
54	36.11	.669	809	2549	0	3358
55	37.55	.683	845	2647	0	3492
56	38.57	.689	864	2723	0	3587
57	38.57	.677	864	2723	0	3587
58	40.46	.698	897	2866	0	3763
59	40.46	.686	897	2866	0	3763
60	40.46	.674	897	2866	0	3763
61	42.44	.696	913	3034	0	3947
62	43.63	.704	926	3132	0	4058
63	43.63	.693	926	3132	0	4058
64	43.63	.682	926	3132	0	4058
65	43.63	.671	926	3132	0	4058
66	43.63	.661	926	3132	0	4058
67	44.90	.670	928	3248	0	4176
68	46.11	.678	954	3334	0	4288
69	47.66	.691	961	3471	0	4432
70	49.76	.711	991	3637	0	4628
71	49.76	.701	991	3637	0	4628
72	50.68	.704	1035	3678	0	4713

Cut-off K	Average Output K'	Overall %Prec-ision	Overall %Known Recall	Average %Prec-ision	Average %Known Recall	S.D. %Prec-ision	S.D. %Known Recall
45	33.15	23.87	35.22	27.64	38.51	22.58	24.50
46	33.15	23.87	35.22	27.64	38.51	22.58	24.50
47	33.15	23.87	35.22	27.64	38.51	22.58	24.50
48	34.86	24.65	38.23	27.30	39.88	21.66	24.55
49	34.86	24.65	38.23	27.30	39.88	21.66	24.55
50	34.86	24.65	38.23	27.30	39.88	21.66	24.55
51	35.47	24.25	38.28	27.28	40.41	21.69	25.31
52	36.11	24.09	38.71	27.22	40.80	21.71	25.27
53	36.11	24.09	38.71	27.22	40.80	21.71	25.27
54	36.11	24.09	38.71	27.22	40.80	21.71	25.27
55	37.55	24.20	40.43	26.85	41.58	21.46	25.35
56	38.57	24.09	41.34	26.92	42.29	21.42	25.20
57	38.57	24.09	41.34	26.92	42.29	21.42	25.20
58	40.46	23.84	42.92	26.66	43.45	21.40	25.46
59	40.46	23.84	42.92	26.66	43.45	21.40	25.46
60	40.46	23.84	42.92	26.66	43.45	21.40	25.46
61	42.44	23.13	43.68	25.89	44.28	21.18	26.02
62	43.63	22.82	44.31	25.84	45.10	21.20	26.00
63	43.63	22.82	44.31	25.84	45.10	21.20	26.00
64	43.63	22.82	44.31	25.84	45.10	21.20	26.00
65	43.63	22.82	44.31	25.84	45.10	21.20	26.00
66	43.63	22.82	44.31	25.84	45.10	21.20	26.00
67	44.90	22.22	44.40	25.56	45.53	21.33	26.24
68	46.11	22.25	45.65	25.35	46.17	21.23	26.12
69	47.66	21.68	45.98	25.05	46.60	21.22	26.16
70	49.76	21.41	47.42	24.96	47.59	21.25	26.15
71	49.76	21.41	47.42	24.96	47.59	21.25	26.15
72	50.68	21.96	49.52	25.07	48.22	21.34	26.06

93 REQUESTS, RUN 28(13W14). TABLE RPK(93,28).

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Not Assessed	Output
0	.00	.000	0	0	0	0
1	.54	.540	18	32	0	50
2	1.41	.705	60	71	0	131
3	2.40	.800	94	129	0	223
4	3.12	.780	114	176	0	290
5	4.13	.826	136	248	0	384
6	5.06	.843	164	307	0	471
7	6.11	.873	185	383	0	568
8	6.86	.858	206	432	0	638
9	8.28	.920	234	536	0	770
10	9.68	.968	278	622	0	900
11	10.80	.982	298	706	0	1004
12	11.54	.962	306	767	0	1073
13	12.55	.965	333	834	0	1167
14	13.13	.938	344	877	0	1221
15	13.83	.922	366	920	0	1286
16	14.52	.908	377	973	0	1350
17	15.39	.905	394	1037	0	1431
18	16.31	.906	432	1085	0	1517
19	17.49	.921	472	1155	0	1627
20	18.63	.932	498	1235	0	1733
21	18.91	.900	499	1260	0	1759
22	20.24	.920	526	1356	0	1882
23	21.54	.937	541	1462	0	2003
24	22.68	.945	549	1560	0	2109
25	23.81	.952	562	1652	0	2214
26	24.96	.960	586	1735	0	2321
27	25.44	.942	597	1769	0	2366
28	27.01	.965	608	1904	0	2512
29	27.58	.951	618	1947	0	2565
30	28.53	.951	625	2028	0	2653
31	29.11	.939	629	2078	0	2707
32	30.29	.947	657	2160	0	2817
33	30.74	.932	666	2193	0	2859
34	31.49	.926	677	2252	0	2929
35	32.13	.918	687	2301	0	2988
36	32.80	.911	692	2358	0	3050
37	34.20	.924	714	2467	0	3181
38	35.05	.922	741	2519	0	3260
39	35.96	.922	747	2597	0	3344
40	37.47	.937	816	2669	0	3485
41	38.42	.937	842	2731	0	3573
42	39.95	.951	860	2855	0	3715
43	41.42	.963	874	2971	7	3852
44	42.46	.965	877	3065	7	3949
45	43.34	.963	896	3128	7	4031
46	44.19	.961	908	3195	7	4110
47	45.35	.965	917	3271	30	4218
48	45.99	.958	920	3327	30	4277
49	46.94	.958	930	3405	30	4365
50	48.37	.967	939	3529	30	4498
51	49.31	.967	949	3601	36	4586
52	50.32	.968	991	3653	36	4680
53	51.12	.965	996	3706	52	4754
54	52.12	.965	1001	3794	52	4847
55	53.22	.968	1019	3878	52	4949
56	53.77	.960	1025	3924	52	5001
57	54.78	.961	1027	4012	56	5095
58	55.56	.958	1036	4073	58	5167
59	56.45	.957	1040	4150	60	5250
60	58.76	.979	1057	4348	60	5465
61	59.58	.977	1061	4420	60	5541
62	60.29	.972	1075	4471	61	5607
63	62.40	.990	1102	4638	63	5803
64	63.45	.991	1110	4717	74	5901
65	64.67	.995	1112	4816	86	6014
66	65.10	.986	1114	4854	86	6054
67	66.10	.987	1128	4933	86	6147
68	66.42	.977	1131	4957	89	6177
69	67.65	.980	1137	5025	129	6291
70	68.73	.982	1146	5107	139	6392
71	70.08	.987	1163	5187	167	6517
72	70.71	.982	1168	5236	172	6576

Cut-off K	Average Output K'	Overall %Prec-ision	Overall %Known Recall	Average %Prec-ision	Average %Known Recall	S.D. %Prec-ision	S.D. %Known Recall
0	.00	.00	.00	.00	.00	.00	.00
1	.54	36.00	.86	19.35	1.02	39.51	2.61
2	1.41	45.80	2.87	33.33	4.32	41.04	7.52
3	2.40	42.15	4.50	37.22	6.92	36.92	9.81
4	3.12	39.31	5.45	38.23	8.15	35.01	10.74
5	4.13	35.42	6.51	36.35	9.58	32.89	11.71
6	5.06	34.82	7.85	37.10	12.33	31.11	15.43
7	6.11	32.57	8.85	33.33	13.28	27.65	15.72
8	6.86	32.29	9.86	33.09	14.47	27.64	16.25
9	8.28	30.39	11.20	31.04	15.96	25.73	16.63
10	9.68	30.89	13.30	32.24	18.90	26.29	19.16
11	10.80	29.68	14.26	31.62	20.10	25.55	19.22
12	11.54	28.52	14.64	30.97	20.91	24.66	19.04
13	12.55	28.53	15.93	30.22	22.33	23.15	19.42
14	13.13	28.17	16.46	29.86	22.75	23.31	19.43
15	13.83	28.46	17.51	29.48	23.73	22.88	19.79
16	14.52	27.93	18.04	29.38	24.51	22.80	19.92
17	15.39	27.53	18.85	28.89	25.20	22.10	19.86
18	16.31	28.48	20.67	28.90	26.04	22.53	19.94
19	17.49	29.01	22.58	28.92	27.46	22.45	19.97
20	18.63	28.74	23.83	28.91	28.68	22.15	19.52
21	18.91	28.37	23.88	28.59	28.75	21.96	19.50
22	20.24	27.95	25.17	28.00	29.89	21.73	20.17
23	21.54	27.01	25.89	27.41	30.72	21.57	20.00
24	22.68	26.03	26.27	26.74	31.04	21.61	19.99
25	23.81	25.38	26.89	26.22	31.59	21.73	20.06
26	24.96	25.25	28.04	26.08	32.50	21.08	19.73
27	25.44	25.23	28.56	25.96	32.89	21.16	19.87
28	27.01	24.20	29.09	25.64	33.53	21.04	20.75
29	27.58	24.09	29.57	25.60	33.88	20.92	20.57
30	28.53	23.56	29.90	25.24	34.35	20.88	20.85
31	29.11	23.24	30.10	25.06	34.53	20.95	20.83
32	30.29	23.32	31.44	24.97	36.02	20.84	21.47
33	30.74	23.29	31.87	24.92	36.41	20.81	21.55
34	31.49	23.11	32.39	24.79	36.86	20.78	21.38
35	32.13	22.99	32.87	24.79	37.35	20.82	21.90
36	32.80	22.69	33.11	24.54	37.54	20.83	21.90
37	34.20	22.45	34.16	23.47	38.26	19.15	21.78
38	35.05	22.73	35.45	23.56	39.56	19.05	22.39
39	35.96	22.34	35.74	23.03	39.75	18.53	22.38
40	37.47	23.41	39.04	23.33	41.13	19.20	22.51
41	38.42	23.57	40.29	23.45	42.05	19.21	22.17
42	39.95	23.15	41.15	23.22	43.10	19.25	22.43
43	41.42	22.73	41.82	22.88	43.98	19.27	22.96
44	42.46	22.25	41.96	22.74	44.07	19.37	22.94
45	43.34	22.27	42.87	22.76	44.76	19.38	22.99
46	44.19	22.13	43.44	22.70	45.23	19.34	22.75
47	45.35	21.90	43.88	22.61	45.99	19.30	22.48
48	45.99	21.66	44.02	22.44	46.16	19.17	22.45
49	46.94	21.45	44.50	22.22	46.66	19.11	22.59
50	48.37	21.02	44.93	21.68	47.00	18.70	22.34
51	49.31	20.86	45.41	21.38	47.31	18.41	22.34
52	50.32	21.34	47.42	21.67	48.83	18.84	22.92
53	51.12	21.18	47.66	21.55	49.09	18.78	23.03
54	52.12	20.88	47.89	21.37	49.48	18.80	23.23
55	53.22	20.81	48.76	20.71	50.11	17.26	23.68
56	53.77	20.71	49.04	20.58	50.26	17.19	23.70
57	54.78	20.38	49.14	20.48	50.80	17.09	23.11
58	55.56	20.28	49.57	20.47	51.45	17.11	23.45
59	56.45	20.04	49.76	20.34	51.54	17.14	23.42
60	58.76	19.56	50.57	19.97	52.66	17.00	23.41
61	59.58	19.36	50.77	19.74	52.79	16.94	23.50
62	60.29	19.38	51.44	19.81	53.35	17.02	23.45
63	62.40	19.20	52.73	19.12	54.06	15.63	22.96
64	63.45	19.05	53.11	18.99	54.26	15.58	22.87
65	64.67	18.76	53.21	18.78	54.32	15.48	22.83
66	65.10	18.67	53.30	18.69	54.40	15.44	22.84
67	66.10	18.61	53.97	18.65	55.30	15.29	22.76
68	66.42	18.58	54.11	18.63	55.44.	15.28	22.74
69	67.65	18.45	54.40	18.54	55.79	15.32	22.74
70	68.73	18.33	54.83	18.49	56.08	15.37	22.77
71	70.08	18.31	55.65	18.33	56.58	15.17	22.67
72	70.71	18.24	55.89	18.29	56.74	15.22	22.68

93 REQUESTS, RUN 14(MCSO1). TABLE RPK(93,14).

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Not Assessed	Output
0	.00	.000	0	0	0	0
1	.40	.400	11	26	0	37
2	1.34	.670	44	81	0	125
3	2.23	.743	74	133	0	207
4	3.25	.813	98	204	0	302
5	4.15	.830	105	281	0	386
6	4.96	.827	129	332	0	461
7	5.65	.807	145	380	0	525
8	6.69	.836	170	452	0	622
9	7.62	.847	189	520	0	709
10	8.74	.874	228	585	0	813
11	9.39	.854	237	636	0	873
12	10.58	.882	260	724	0	984
13	11.18	.860	266	774	0	1040
14	12.05	.861	281	840	0	1121
15	13.34	.889	296	945	0	1241
16	14.44	.903	308	1035	0	1343
17	14.90	.876	314	1072	0	1386
18	15.75	.875	321	1144	0	1465
19	16.98	.894	332	1247	0	1579
20	18.37	.919	374	1334	0	1708
21	19.15	.912	383	1398	0	1781
22	19.67	.894	394	1435	0	1829
23	21.68	.943	428	1588	0	2016
24	22.42	.934	450	1635	0	2085
25	23.62	.945	457	1740	0	2197
26	23.95	.921	460	1767	0	2227
27	25.18	.933	487	1855	0	2342
28	26.48	.946	504	1959	0	2463
29	27.24	.939	513	2020	0	2533
30	27.92	.931	527	2070	0	2597
31	28.44	.917	527	2118	0	2645
32	28.70	.897	531	2138	0	2669
33	29.69	.900	546	2215	0	2761
34	29.83	.877	548	2226	0	2774
35	30.03	.858	549	2244	0	2793
36	30.63	.851	559	2290	0	2849
37	32.28	.872	579	2423	0	3002
38	33.90	.892	606	2547	0	3153
39	35.31	.905	647	2637	0	3284
40	36.73	.918	708	2708	0	3416
41	37.49	.914	713	2774	0	3487
42	38.30	.912	728	2834	0	3562
43	39.74	.924	753	2943	0	3696
44	40.95	.931	755	3053	0	3808
45	41.47	.922	760	3097	0	3857
46	43.03	.935	774	3228	0	4002
47	43.61	.928	791	3265	0	4056
48	44.52	.928	799	3341	0	4140
49	45.15	.921	812	3387	0	4199
50	46.30	.926	829	3477	0	4306
51	46.89	.919	833	3528	0	4361
52	47.10	.906	834	3546	0	4380
53	48.25	.910	844	3643	0	4487
54	48.87	.905	849	3696	0	4545
55	51.00	.927	873	3870	0	4743

Cut-off K	Average Output K′	Overall %Precision	Overall %Known Recall	Average %Precision	Average %Known Recall	S.D. %Precision	S.D. %Known Recall
0	.00	.00	.00	.00	.00	.00	.00
1	.40	29.73	.53	11.83	.58	32.29	1.92
2	1.34	35.20	2.11	23.66	2.92	37.17	6.55
3	2.23	35.75	3.54	27.78	4.64	35.33	8.00
4	3.25	32.45	4.69	30.40	5.91	34.54	8.55
5	4.15	27.20	5.02	28.31	6.12	33.20	8.60
6	4.96	27.98	6.17	28.37	7.37	31.01	9.17
7	5.65	27.62	6.94	29.36	7.92	31.05	9.37
8	6.69	27.33	8.13	29.18	9.83	29.69	10.91
9	7.62	26.66	9.04	27.84	10.50	27.50	11.32
10	8.74	28.04	10.91	28.66	13.15	26.32	15.75
11	9.39	27.15	11.34	28.63	13.63	28.13	15.80
12	10.58	26.42	12.44	28.52	14.95	27.71	16.13
13	11.18	25.58	12.73	27.47	15.21	26.53	16.16
14	12.05	25.07	13.44	27.28	16.66	26.51	18.29
15	13.34	23.85	14.16	26.83	17.48	26.18	18.35
16	14.44	22.93	14.74	25.94	17.84	25.72	18.29
17	14.90	22.66	15.02	25.76	18.17	25.51	18.53
18	15.75	21.91	15.36	25.06	18.79	25.16	18.93
19	16.98	21.03	15.89	24.01	19.13	24.51	18.92
20	18.37	21.90	17.89	23.46	20.33	23.96	19.53
21	19.15	21.50	18.33	23.13	21.17	23.25	20.13
22	19.67	21.54	18.85	23.14	21.70	23.07	19.97
23	21.63	21.23	20.48	22.65	23.26	22.86	20.60
24	22.42	21.58	21.53	22.41	24.08	22.67	21.41
25	23.62	20.80	21.87	21.69	24.38	21.68	21.38
26	23.95	20.66	22.01	21.63	24.51	21.68	21.32
27	25.18	20.79	23.30	21.93	26.01	21.37	21.24
28	26.48	20.46	24.11	21.72	26.67	21.24	21.20
29	27.24	20.25	24.55	21.58	27.13	21.20	21.46
30	27.92	20.29	25.22	21.64	27.46	21.27	21.30
31	28.44	19.92	25.22	21.61	27.46	21.28	21.30
32	28.70	19.90	25.41	21.54	27.67	21.22	21.48
33	29.69	19.78	26.12	21.36	28.33	21.21	21.74
34	29.83	19.75	26.22	21.37	28.49	21.20	21.76
35	30.03	19.66	26.27	21.31	28.54	21.21	21.75
36	30.63	19.62	26.75	21.23	28.77	21.07	21.66
37	32.28	19.29	27.70	20.22	29.74	19.49	21.81
38	33.90	19.22	29.00	20.16	30.85	19.43	22.12
39	35.31	19.70	30.96	20.53	31.44	20.15	22.12
40	36.73	20.73	33.88	20.56	32.52	20.63	22.62
41	37.49	20.45	34.11	20.43	32.77	20.62	22.57
42	38.30	20.44	34.83	20.44	33.31	20.57	22.50
43	39.74	20.37	36.03	20.57	34.11	20.51	22.10
44	40.95	19.83	36.12	20.35	34.27	20.50	22.10
45	41.47	19.70	36.36	20.32	34.60	20.49	22.40
46	43.03	19.34	37.03	19.97	35.07	20.41	22.41
47	43.61	19.50	37.85	19.80	35.61	20.07	22.89
48	44.52	19.30	38.23	19.89	36.70	20.00	22.89
49	45.15	19.34	38.85	19.91	37.21	19.98	22.90
50	46.30	19.25	39.67	19.72	37.76	19.91	22.95
51	46.89	19.10	39.86	19.74	38.41	19.88	22.81
52	47.10	19.04	39.90	19.68	38.42	19.87	22.81
53	48.25	18.81	40.38	19.43	38.81	19.83	22.72
54	48.87	18.68	40.62	19.38	39.07	19.76	22.68
55	51.00	18.41	41.77	18.63	40.40	18.09	22.94

55 UNDERLINED REQUESTS, RUN 25(U13). TABLE RPK(55,25).

Cut-off K	Average Output K'	K'/K	Relevant	Irrelevant	Not Assessed	Output
0	.00	.000	0	0	0	0
1	.31	.310	10	7	0	17
2	.98	.490	37	17	0	54
3	1.98	.660	59	50	0	109
4	2.55	.638	75	65	0	140
5	3.04	.608	83	84	0	167
6	4.16	.693	126	103	0	229
7	4.84	.691	144	122	0	266
8	5.18	.648	156	129	0	285
9	5.82	.647	167	153	0	320
10	6.85	.685	194	183	0	377
11	7.51	.683	203	210	0	413
12	8.36	.697	226	234	0	460
13	8.78	.675	230	253	0	483
14	9.47	.676	238	283	0	521
15	9.47	.631	238	283	0	521
16	10.33	.646	240	328	0	568
17	10.33	.608	240	328	0	568
18	11.11	.617	265	346	0	611
19	11.78	.620	287	361	0	648
20	12.78	.639	311	392	0	703
21	14.04	.669	317	455	0	772
22	14.58	.663	327	475	0	802
23	15.62	.679	336	523	0	859
24	15.62	.651	336	523	0	859
25	15.69	.628	337	526	0	863
26	16.02	.616	340	541	0	881
27	17.45	.646	355	605	0	960
28	18.49	.660	356	660	1	1017
29	18.49	.638	356	660	1	1017
30	18.49	.616	356	660	1	1017
31	18.49	.596	356	660	1	1017
32	19.24	.601	378	679	1	1058
33	19.73	.598	382	697	6	1085
34	21.24	.625	399	763	6	1168
35	23.11	.660	441	824	6	1271
36	24.25	.674	445	883	6	1334
37	25.04	.677	445	924	8	1377
38	26.27	.691	453	984	8	1445
39	26.27	.674	453	984	8	1445
40	27.53	.688	462	1044	8	1514
41	28.55	.696	470	1092	8	1570
42	29.49	.702	478	1130	14	1622
43	29.49	.686	478	1130	14	1622
44	29.49	.670	478	1130	14	1622

Cut-off K	Average Output K'	Overall %Prec-ision	Cverall %Known Recall	Average %Prec-ision	Average %Known Recall	S.D. %Prec-ision	S.D. %Known Recall
0	.00	.00	.00	.00	.00	.00	.00
1	.31	58.82	.71	18.18	1.92	38.57	7.10
2	.98	68.52	2.64	34.24	4.25	44.09	8.58
3	1.98	54.13	4.21	39.76	6.97	41.50	10.44
4	2.55	53.57	5.35	44.27	7.98	40.75	10.38
5	3.04	49.70	5.92	43.97	8.77	39.25	11.14
6	4.16	55.02	8.99	48.77	10.96	38.41	11.74
7	4.84	54.14	10.28	51.50	12.11	37.54	12.17
8	5.18	54.74	11.13	50.73	13.42	37.02	13.98
9	5.82	52.19	11.92	50.22	14.48	36.12	13.93
10	6.85	51.46	13.85	50.07	16.67	35.98	16.15
11	7.51	49.15	14.49	49.07	17.46	35.95	16.13
12	8.36	49.13	16.13	48.31	19.13	35.46	17.03
13	8.78	47.62	16.42	48.61	19.40	35.11	16.84
14	9.47	45.68	16.99	47.66	20.04	34.41	17.24
15	9.47	45.68	16.99	47.66	20.04	34.41	17.24
16	10.33	42.25	17.13	46.10	20.52	34.83	17.56
17	10.33	42.25	17.13	46.10	20.52	34.83	17.56
18	11.11	43.37	18.92	47.18	22.04	34.36	19.22
19	11.78	44.29	20.49	47.06	22.87	34.29	19.29
20	12.78	44.24	22.20	45.41	23.85	32.95	19.40
21	14.04	41.06	22.63	44.03	24.80	33.53	19.84
22	14.58	40.77	23.34	43.53	25.45	33.39	20.07
23	15.62	39.12	23.98	42.82	26.25	33.50	20.26
24	15.62	39.12	23.98	42.82	26.25	33.50	20.26
25	15.69	39.05	24.05	42.78	26.42	33.50	20.75
26	16.02	38.59	24.27	42.64	26.69	33.57	20.82
27	17.45	36.98	25.34	42.79	28.85	33.31	22.44
28	18.49	35.04	25.41	42.57	28.98	33.46	22.46
29	18.49	35.04	25.41	42.57	28.98	33.46	22.46
30	18.49	35.04	25.41	42.57	28.98	33.46	22.46
31	18.49	35.04	25.41	42.57	28.98	33.46	22.46
32	19.24	35.76	26.98	42.43	29.84	33.39	22.83
33	19.73	35.40	27.27	41.89	30.14	33.12	23.26
34	21.24	34.34	28.48	40.86	31.62	32.14	23.91
35	23.11	34.86	31.48	40.55	33.68	31.11	24.00
36	24.25	33.51	31.76	40.24	34.59	31.36	24.13
37	25.04	32.51	31.76	39.79	34.59	31.62	24.13
38	26.27	31.52	32.33	38.26	35.80	30.70	25.27
39	26.27	31.52	32.33	38.26	35.80	30.70	25.27
40	27.53	30.68	32.98	38.14	36.76	30.78	25.12
41	28.55	30.09	33.55	37.10	37.28	30.20	25.38
42	29.49	29.73	34.12	36.79	37.72	30.25	25.26
43	29.49	29.73	34.12	36.79	37.72	30.25	25.26
44	29.49	29.73	34.12	36.79	37.72	30.25	25.26

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Not Assessed	Output
45	29.49	.655	478	1130	14	1622
46	29.49	.641	478	1130	14	1622
47	29.49	.627	478	1130	14	1622
48	29.49	.614	478	1130	14	1622
49	29.93	.611	485	1147	14	1646
50	30.47	.609	489	1169	18	1676
51	30.47	.597	489	1169	18	1676
52	30.47	.586	489	1169	18	1676
53	30.47	.575	489	1169	18	1676
54	31.78	.589	496	1203	49	1748
55	31.78	.578	496	1203	49	1748
56	31.78	.568	496	1203	49	1748
57	31.78	.558	496	1203	49	1748
58	33.69	.581	511	1250	92	1853
59	33.69	.571	511	1250	92	1853
60	33.69	.562	511	1250	92	1853
61	35.15	.576	519	1268	146	1933
62	37.49	.605	542	1327	193	2062
63	38.78	.616	542	1361	230	2133
64	38.78	.606	542	1361	230	2133
65	38.78	.597	542	1361	230	2133
66	40.67	.616	544	1413	280	2237
67	40.67	.607	544	1413	280	2237
68	42.78	.629	545	1421	387	2353
69	42.78	.620	545	1421	387	2353
70	43.89	.627	549	1456	409	2414
71	44.27	.624	549	1458	428	2435
72	44.27	.615	519	1458	428	2435
73	46.25	.634	564	1494	486	2544
74	46.25	.625	564	1494	486	2544
75	48.29	.644	578	1557	521	2656
76	48.29	.635	578	1557	521	2656
77	48.29	.627	578	1557	521	2656
78	48.29	.619	578	1557	521	2656
79	48.82	.618	579	1558	548	2685
80	48.82	.610	579	1558	548	2685
81	50.40	.622	579	1589	604	2772
82	50.40	.615	579	1589	604	2772

Cut-off K	Average Output K'	Overall %Prec-ision	Overall %Known Recall	Average %Prec-ision	Average %Known Recall	S.D. %Prec-ision	S.D. %Known Recall
45	29.49	29.73	34.12	36.79	37.72	30.25	25.26
46	29.49	29.73	34.12	36.79	37.72	30.25	25.26
47	29.49	29.73	34.12	36.79	37.72	30.25	25.26
48	29.49	29.73	34.12	36.79	37.72	30.25	25.26
49	29.93	29.72	34.62	36.72	38.18	30.25	25.69
50	30.47	29.49	34.90	36.66	38.54	30.29	25.86
51	30.47	29.49	34.90	36.66	38.54	30.29	25.86
52	30.47	29.49	34.90	36.66	38.54	30.29	25.86
53	30.47	29.49	34.90	36.66	38.54	30.29	25.86
54	31.78	29.19	35.40	36.42	38.93	30.34	25.71
55	31.78	29.19	35.40	36.42	38.93	30.34	25.71
56	31.78	29.19	35.40	36.42	38.93	30.34	25.71
57	31.78	29.19	35.40	36.42	38.93	30.34	25.71
58	33.69	29.02	36.47	36.01	39.45	30.27	25.62
59	33.69	29.02	36.47	36.01	39.45	30.27	25.62
60	33.69	29.02	36.47	36.01	39.45	30.27	25.62
61	35.15	29.04	37.04	35.92	40.22	30.26	26.31
62	37.49	29.00	38.69	35.05	41.11	29.63	26.82
63	38.78	28.48	38.69	34.98	41.11	29.71	26.82
64	38.78	28.48	38.69	34.98	41.11	29.71	26.82
65	38.78	28.48	38.69	34.98	41.11	29.71	26.82
66	40.67	27.80	38.83	34.24	41.37	29.73	26.97
67	40.67	27.80	38.83	34.24	41.37	29.73	26.97
68	42.78	27.72	38.90	33.64	41.39	29.17	26.95
69	42.78	27.72	38.90	33.64	41.39	29.17	26.95
70	43.89	27.38	39.19	33.61	41.91	29.19	27.11
71	44.27	27.35	39.19	33.57	41.91	29.14	27.11
72	44.27	27.35	39.19	33.57	41.91	29.14	27.11
73	46.25	27.41	40.26	33.37	42.59	29.10	27.02
74	46.25	27.41	40.26	33.37	42.59	29.10	27.02
75	48.29	27.07	41.26	32.24	43.39	27.89	27.95
76	48.29	27.07	41.26	32.24	43.39	27.89	27.95
77	48.29	27.07	41.26	32.24	43.39	27.89	27.95
78	48.29	27.07	41.26	32.24	43.39	27.89	27.95
79	48.82	27.09	41.33	32.26	43.48	27.88	28.01
80	48.82	27.09	41.33	32.26	43.48	27.88	28.01
81	50.40	26.71	41.33	32.13	43.48	27.98	28.01
82	50.40	26.71	41.33	32.13	43.48	27.98	28.01

In this and the following two tables, documents not assessed in any standard or manual strategy have been set irrelevant.

93 REQUESTS, RUN 13(KWS) CONTINUED. TABLE RPL(93,13).

Cut-off K	Average Output K'	K'/K	Relevant	Irrelevant	Output	Overall %Precision	Overall %Known Recall
73	51.66	.708	1042	3762	4804	21.69	49.86
74	51.66	.698	1042	3762	4804	21.69	49.86
75	51.66	.689	1042	3762	4804	21.69	49.86
76	52.56	.692	1047	3841	4888	21.42	50.10
77	52.56	.683	1047	3841	4888	21.42	50.10
78	52.56	.674	1047	3841	4888	21.42	50.10
79	52.56	.665	1047	3841	4888	21.42	50.10
80	52.56	.657	1047	3841	4888	21.42	50.10
81	52.56	.649	1047	3841	4888	21.42	50.10
82	54.02	.659	1059	3965	5024	21.08	50.67
83	54.02	.651	1059	3965	5024	21.08	50.67
84	54.02	.643	1059	3965	5024	21.08	50.67
85	55.75	.656	1062	4123	5185	20.48	50.81
86	55.75	.648	1062	4123	5185	20.48	50.81
87	57.41	.660	1064	4275	5339	19.93	50.91
88	57.41	.652	1064	4275	5339	19.93	50.91
89	57.41	.645	1064	4275	5339	19.93	50.91
90	58.89	.654	1085	4392	5477	19.81	51.91
91	58.89	.647	1085	4392	5477	19.81	51.91
92	60.47	.657	1100	4524	5624	19.56	52.63
93	60.47	.650	1100	4524	5624	19.56	52.63
94	62.06	.660	1110	4662	5772	19.23	53.11
95	62.06	.653	1110	4662	5772	19.23	53.11
96	62.06	.646	1110	4662	5772	19.23	53.11
97	62.06	.640	1110	4662	5772	19.23	53.11
98	65.04	.664	1144	4905	6049	18.91	54.74
99	65.04	.657	1144	4905	6049	18.91	54.74
100	65.04	.650	1144	4905	6049	18.91	54.74
101	65.04	.644	1144	4905	6049	18.91	54.74
102	65.04	.638	1144	4905	6049	18.91	54.74
103	65.04	.631	1144	4905	6049	18.91	54.74
104	67.10	.645	1158	5082	6240	18.56	55.41
105	67.10	.639	1158	5082	6240	18.56	55.41
106	67.10	.633	1158	5082	6240	18.56	55.41
107	67.10	.627	1158	5082	6240	18.56	55.41
108	67.10	.621	1158	5082	6240	18.56	55.41
109	67.10	.616	1158	5082	6240	18.56	55.41
110	69.27	.630	1166	5276	6442	18.10	55.79
111	71.49	.644	1167	5482	6649	17.55	55.84
112	73.83	.659	1184	5682	6866	17.24	56.65
113	75.97	.672	1185	5880	7065	16.77	56.70
114	75.97	.666	1185	5880	7065	16.77	56.70
115	75.97	.661	1185	5880	7065	16.77	56.70
116	79.45	.685	1204	6185	7389	16.29	57.61
117	79.45	.679	1204	6185	7389	16.29	57.61
118	81.43	.690	1221	6352	7573	16.12	58.42

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Output	Overall %Precision	Overall %Known Recall
119	81.43	.684	1221	6352	7573	16.12	58.42
120	81.43	.679	1221	6352	7573	16.12	58.42
121	81.43	.673	1221	6352	7573	16.12	58.42
122	83.78	.687	1225	6567	7792	15.72	58.61
123	83.78	.681	1225	6567	7792	15.72	58.61
124	83.78	.676	1225	6567	7792	15.72	58.61
125	86.08	.689	1226	6779	8005	15.32	58.66
126	86.08	.683	1226	6779	8005	15.32	58.66
127	86.08	.678	1226	6779	8005	15.32	58.66
128	86.08	.673	1226	6779	8005	15.32	58.66
129	90.76	.704	1259	7182	8441	14.92	60.24
130	92.95	.715	1268	7376	8644	14.67	60.67
131	95.37	.728	1277	7592	8869	14.40	61.10
132	95.37	.723	1277	7592	8869	14.40	61.10
133	97.19	.731	1290	7749	9039	14.27	61.72
134	99.14	.740	1313	7907	9220	14.24	62.82
135	99.14	.734	1313	7907	9220	14.24	62.82
136	99.14	.729	1313	7907	9220	14.24	62.82
137	99.14	.724	1313	7907	9220	14.24	62.82
138	99.14	.718	1313	7907	9220	14.24	62.82
139	103.75	.746	1352	8297	9649	14.01	64.69
140	106.34	.760	1356	8534	9890	13.71	64.88
141	106.34	.754	1356	8534	9890	13.71	64.88
142	106.34	.749	1356	8534	9890	13.71	64.88
143	107.80	.754	1362	8663	10025	13.59	65.17
144	107.80	.749	1362	8663	10025	13.59	65.17
145	107.80	.743	1362	8663	10025	13.59	65.17
146	107.80	.738	1362	8663	10025	13.59	65.17
147	107.80	.733	1362	8663	10025	13.59	65.17
148	107.80	.728	1362	8663	10025	13.59	65.17
149	107.80	.723	1362	8663	10025	13.59	65.17
150	107.80	.719	1362	8663	10025	13.59	65.17
151	110.63	.733	1367	8922	10289	13.29	65.41
152	112.14	.738	1375	9054	10429	13.18	65.79
153	112.14	.733	1375	9054	10429	13.18	65.79
154	115.89	.753	1401	9377	10778	13.00	67.03
155	115.89	.748	1401	9377	10778	13.00	67.03
156	115.89	.743	1401	9377	10778	13.00	67.03
157	115.89	.738	1401	9377	10778	13.00	67.03
158	115.89	.733	1401	9377	10778	13.00	67.03
159	115.89	.729	1401	9377	10778	13.00	67.03
160	121.02	.756	1415	9840	11255	12.57	67.70
161	123.97	.770	1430	10099	11529	12.40	68.42
162	123.97	.765	1430	10099	11529	12.40	68 42
163	126.73	.777	1473	10313	11786	12.50	70.48
164	126.73	.773	1473	10313	11786	12.50	70.48
165	126.73	.768	1473	10313	11786	12.50	70.48
166	126.73	.763	1473	10313	11786	12.50	70.48
167	126.73	.759	1473	10313	11786	12.50	70.48
168	128.82	.767	1474	10506	11980	12.30	70.53
169	128.82	.762	1474	10506	11980	12.30	70.53
170	128.82	.758	1474	10506	11980	12.30	70.53
171	131.01	.766	1478	10706	12184	12.13	70.72
172	131.01	.762	1478	10706	12184	12.13	70.72
173	131.01	.757	1478	10706	12184	12.13	70.72
174	133.34	.766	1485	10916	12401	11.97	71.05
175	133.34	.762	1485	10916	12401	11.97	71.05
176	136.39	.775	1488	11196	12684	11.73	71.20
177	136.39	.771	1488	11196	12684	11.73	71.20
178	136.39	.766	1488	11196	12684	11.73	71.20
179	136.39	.762	1488	11196	12684	11.73	71.20
180	136.39	.758	1488	11196	12684	11.73	71.20
181	136.39	.754	1488	11196	12684	11.73	71.20
182	136.39	.749	1488	11196	12684	11.73	71.20

Cut-off K	Average Output K′	K′/K	Rele-vant	Irrel-evant	Output	Overall %Prec-ision	Overall %Known Recall
183	139.63	.763	1491	11495	12986	11.48	71.34
184	139.63	.759	1491	11495	12986	11.48	71.34
185	139.63	.755	1491	11495	12986	11.48	71.34
186	139.63	.751	1491	11495	12986	11.48	71.34
187	139.63	.747	1491	11495	12986	11.48	71.34
188	139.63	.743	1491	11495	12986	11.48	71.34
189	142.56	.754	1510	11748	13258	11.39	72.25
190	146.23	.770	1517	12082	13599	11.16	72.58
191	146.23	.766	1517	12082	13599	11.16	72.58
192	149.40	.778	1531	12363	13894	11.02	73.25
193	149.40	.774	1531	12363	13894	11.02	73.25
194	153.30	.790	1534	12723	14257	10.76	73.40
195	156.46	.802	1537	13014	14551	10.56	73.54
196	156.46	.798	1537	13014	14551	10.56	73.54
197	156.46	.794	1537	13014	14551	10.56	73.54
198	160.19	.809	1543	13355	14898	10.36	73.83
199	163.81	.823	1555	13679	15234	10.21	74.40
200	163.81	.819	1555	13679	15234	10.21	74.40
210	169.53	.807	1565	14201	15766	9.93	74.88
220	183.20	.833	1592	15446	17038	9.34	76.17
230	190.43	.828	1606	16104	17710	9.07	76.84
240	199.99	.833	1632	16967	18599	8.77	78.09
250	206.68	.827	1637	17584	19221	8.52	78.33
260	217.72	.837	1663	18585	20248	8.21	79.57
270	224.67	.832	1675	19219	20894	8.02	80.14
280	229.53	.820	1676	19670	21346	7.85	80.19
290	229.53	.791	1676	19670	21346	7.85	80.19
300	234.43	.781	1679	20123	21802	7.70	80.33
310	250.28	.807	1687	21589	23276	7.25	80.72
320	260.65	.815	1690	22550	24240	6.97	80.86
330	266.01	.806	1696	23043	24739	6.86	81.15
340	270.99	.797	1704	23498	25202	6.76	81.53
350	270.99	.774	1704	23498	25202	6.76	81.53
360	276.41	.768	1713	23993	25706	6.66	81.96
370	281.54	.761	1768	24415	26183	6.75	84.59
380	287.67	.757	1787	24966	26753	6.68	85.50
390	287.67	.738	1787	24966	26753	6.68	85.50
400	295.04	.738	1791	25648	27439	6.53	85.69
410	295.04	.720	1791	25648	27439	6.53	85.69
420	306.63	.730	1796	26721	28517	6.30	85.93
430	306.63	.713	1796	26721	28517	6.30	85.93
440	306.63	.697	1796	26721	28517	6.30	85.93
450	331.73	.737	1834	29017	30851	5.94	87.75
460	345.68	.751	1846	30302	32148	5.74	88.33
470	365.96	.779	1862	32172	34034	5.47	89.09
480	365.96	.762	1862	32172	34034	5.47	89.09
490	365.96	.747	1862	32172	34034	5.47	89.09
500	365.96	.732	1862	32172	34034	5.47	89.09

93 REQUESTS, RUN 14(MCSO1) CONTINUED. TABLE RPL(93,14).

Cut-off K	Average Output K′	K′/K	Rele-vant	Irrel-evant	Output	Overall % Prec-ision	Overall % Known Recall
55	51.00	.927	873	3870	4743	18.41	41.77
56	52.74	.942	876	4029	4905	17.86	41.91
57	53.76	.943	893	4107	5000	17.86	42.73
58	54.25	.935	895	4150	5045	17.74	42.82
59	55.30	.937	911	4232	5143	17.71	43.59
60	55.54	.926	912	4253	5165	17.66	43.64
61	56.11	.920	920	4298	5218	17.63	44.02
62	57.41	.926	939	4400	5339	17.59	44.93
63	57.43	.912	939	4402	5341	17.58	44.93
64	57.43	.897	939	4402	5341	17.58	44.93
65	58.86	.906	944	4530	5474	17.25	45.17
66	59.15	.896	945	4556	5501	17.18	45.22
67	59.18	.883	946	4558	5504	17.19	45.26
68	59.90	.881	951	4620	5571	17.07	45.50
69	60.43	.876	953	4667	5620	16.96	45.60
70	61.82	.883	966	4783	5749	16.80	46.22
71	62.30	.877	966	4828	5794	16.67	46.22
72	64.25	.892	978	4997	5975	16.37	46.79
73	65.32	.895	989	5086	6075	16.28	47.32
74	65.96	.891	991	5143	6134	16.16	47.42
75	66.19	.883	995	5161	6156	16.16	47.61
76	66.91	.880	1000	5223	6223	16.07	47.85
77	68.91	.895	1010	5399	6409	15.76	48.33
78	69.60	.892	1011	5462	6473	15.62	48.37
79	71.39	.904	1036	5603	6639	15.60	49.57
80	73.75	.922	1041	5818	6859	15.18	49.81
81	75.14	.928	1044	5944	6988	14.94	49.95
82	76.68	.935	1047	6084	7131	14.68	50.10
83	77.13	.929	1053	6120	7173	14.68	50.38
84	79.61	.948	1059	6345	7404	14.30	50.67
85	80.37	.946	1061	6413	7474	14.20	50.77
86	81.48	.947	1067	6511	7578	14.08	51.05
87	82.97	.954	1070	6646	7716	13.87	51.20
88	84.73	.963	1079	6801	7880	13.69	51.63
89	85.00	.955	1080	6825	7905	13.66	51.67
90	85.86	.954	1080	6905	7985	13.53	51.67
91	86.57	.951	1089	6962	8051	13.53	52.11
92	86.84	.944	1089	6987	8076	13.48	52.11
93	88.10	.947	1097	7096	8193	13.39	52.49
94	89.90	.956	1099	7262	8361	13.14	52.58
95	90.39	.951	1099	7307	8406	13.07	52.58
96	91.43	.952	1102	7401	8503	12.96	52.73
97	93.18	.961	1105	7561	8666	12.75	52.87
98	94.66	.966	1111	7692	8803	12.62	53.16
99	95.22	.962	1111	7744	8855	12.55	53.16
100	95.85	.959	1119	7795	8914	12.55	53.54
101	96.91	.960	1137	7876	9013	12.62	54.40
102	97.65	.957	1141	7940	9081	12.56	54.59
103	98.05	.952	1141	7978	9119	12.51	54.59
104	98.95	.951	1145	8057	9202	12.44	54.78
105	100.20	.954	1146	8173	9319	12.30	54.83
106	100.45	.948	1147	8195	9342	12.28	54.88
107	102.60	.959	1153	8389	9542	12.08	55.17
108	102.97	.953	1153	8423	9576	12.04	55.17
109	103.87	.953	1158	8502	9660	11.99	55.41
110	104.12	.947	1160	8523	9683	11.98	55.50
111	104.43	.941	1161	8551	9712	11.95	55.55
112	104.72	.935	1161	8578	9739	11.92	55.55
113	106.14	.939	1162	8709	9871	11.77	55.60
114	106.28	.932	1162	8722	9884	11.76	55.60
115	106.40	.925	1163	8732	9895	11.75	55.65
116	106.40	.917	1163	8732	9895	11.75	55.65
117	106.95	.914	1163	8783	9946	11.69	55.65
118	107.53	.911	1165	8835	10000	11.65	55.74

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Output	Overall % Precision	Overall % Known Recall
119	109.60	.921	1170	9023	10193	11.48	55.98
120	110.28	.919	1171	9085	10256	11.42	56.03
121	110.85	.916	1173	9136	10309	11.38	56.12
122	111.18	.911	1175	9165	10340	11.36	56.22
123	113.00	.919	1185	9324	10509	11.28	56.70
124	113.11	.912	1185	9334	10519	11.27	56.70
125	114.11	.913	1187	9425	10612	11.19	56.79
126	115.39	.916	1188	9543	10731	11.07	56.84
127	115.39	.909	1188	9543	10731	11.07	56.84
128	115.40	.902	1188	9544	10732	11.07	56.84
129	115.82	.898	1190	9581	10771	11.05	56.94
130	115.82	.891	1190	9581	10771	11.05	56.94
131	117.73	.899	1194	9755	10949	10.91	57.13
132	119.97	.909	1206	9951	11157	10.81	57.70
133	122.08	.918	121[illegible]	10143	11353	10.66	57.89
134	122.92	.917	121[illegible]	10222	11432	10.58	57.89
135	124.20	.920	12[illegible]	10335	11551	10.53	58.18
136	125.77	.925	1217	10480	11697	10.40	58.23
137	128.88	.941	1232	10754	11986	10.28	58.95
138	130.97	.949	1233	10947	12180	10.12	59.00
139	131.15	.944	1233	10964	12197	10.11	59.00
140	131.75	.941	1255	10998	12253	10.24	60.05
141	134.97	.957	1256	11296	12552	10.01	60.10
142	135.67	.955	1256	11361	12617	9.95	60.10
143	136.73	.956	1258	11458	12716	9.89	60.19
144	137.16	.953	1264	11492	12756	9.91	60.48
145	138.11	.952	1272	11572	12844	9.90	60.86
146	138.28	.947	1275	11585	12860	9.91	61.00
147	138.69	.943	1275	11623	12898	9.89	61.00
148	139.43	.942	1277	11690	12967	9.85	61.10
149	141.00	.946	1277	11836	13113	9.74	61.10
150	142.87	.952	1286	12001	13287	9.68	61.53
151	143.57	.951	1287	12065	13352	9.64	61.58
152	143.57	.945	1287	12065	13352	9.64	61.58
153	143.57	.938	1287	12065	13352	9.64	61.58
154	144.05	.935	1287	12110	13397	9.61	61.58
155	144.82	.934	1289	12179	13468	9.57	61.67
156	144.82	.928	1289	12179	13468	9.57	61.67
157	146.39	.932	1292	12322	13614	9.49	61.82
158	147.02	.931	1292	12381	13673	9.45	61.82
159	147.02	.925	1292	12381	13673	9.45	61.82
160	147.02	.919	1292	12381	13673	9.45	61.82
161	149.98	.932	1296	12652	13948	9.29	62.01
162	152.01	.938	1299	12838	14137	9.19	62.15
163	152.01	.933	1299	12838	14137	9.19	62.15
164	152.86	.932	1299	12917	14216	9.14	62.15
165	153.75	.932	1299	13000	14299	9.08	62.15
166	155.57	.937	1302	13166	14468	9.00	62.30
167	155.70	.932	1302	13178	14480	8.99	62.30
168	156.88	.934	1310	13280	14590	8.98	62.68
169	160.27	.948	1313	13592	14905	8.81	62.82
170	164.17	.966	1314	13954	15268	8.61	62.87
171	165.74	.969	1314	14100	15414	8.52	62.87
172	166.78	.970	1316	14195	15511	8.48	62.97
173	166.78	.964	1316	14195	15511	8.48	62.97
174	168.68	.969	1326	14361	15687	8.45	63.44
175	171.19	.978	1327	14594	15921	8.33	63.49
176	171.60	.975	1329	14630	15959	8.33	63.59
177	175.23	.990	1331	14965	16296	8.17	63.68
178	176.03	.989	1333	15038	16371	8.14	63.78
179	176.03	.983	1333	15038	16371	8.14	63.78
180	177.53	.986	1335	15175	16510	8.09	63.88
181	177.53	.981	1335	15175	16510	8.09	63.88
182	177.53	.975	1335	15175	16510	8.09	63.88

Cut-off K	Average Output K'	K'/K	Relevant	Irrelevant	Output	Overall % Precision	Overall % Known Recall
183	177.53	.970	1335	15175	16510	8.09	63.88
184	179.90	.978	1337	15394	16731	7.99	63.97
185	180.14	.974	1337	15416	16753	7.98	63.97
186	180.91	.973	1341	15484	16825	7.97	64.16
187	183.38	.981	1355	15699	17054	7.95	64.83
188	185.04	.984	1356	15853	17209	7.88	64.88
189	185.47	.981	1357	15892	17249	7.87	64.93
190	185.73	.978	1357	15916	17273	7.86	64.93
191	187.47	.982	1358	16077	17435	7.79	64.98
192	189.01	.984	1365	16213	17578	7.77	65.31
193	189.74	.983	1367	16279	17646	7.75	65.41
194	189.74	.978	1367	16279	17646	7.75	65.41
195	193.92	.994	1368	16667	18035	7.59	65.45
196	194.00	.990	1368	16674	18042	7.58	65.45
197	196.15	.996	1373	16869	18242	7.53	65.69
198	196.52	.993	1375	16901	18276	7.52	65.79
199	196.52	.988	1375	16901	18276	7.52	65.79
200	197.31	.987	1377	16973	18350	7.50	65.89
210	204.38	.973	1391	17616	19007	7.32	66.56
220	216.24	.983	1403	18707	20110	6.98	67.13
230	228.29	.993	1424	19807	21231	6.71	68.13
240	234.24	.976	1437	20347	21784	6.60	68.76
250	242.77	.971	1450	21128	22578	6.42	69.38
260	255.15	.981	1455	22274	23729	6.13	69.62
270	263.59	.976	1468	23046	24514	5.99	70.24
280	275.69	.985	1480	24159	25639	5.77	70.81
290	284.68	.982	1489	24986	26475	5.62	71.24
300	291.96	.973	1494	25658	27152	5.50	71.48
310	300.75	.970	1501	26469	27970	5.37	71.82
320	305.84	.956	1508	26935	28443	5.30	72.15
330	316.05	.958	1524	27869	29393	5.18	72.92
340	320.73	.943	1534	28294	29828	5.14	73.40
350	328.16	.938	1544	28975	30519	5.06	73.88
360	340.55	.946	1549	30122	31671	4.89	74.11
370	345.16	.933	1554	30546	32100	4.84	74.35
380	363.46	.956	1559	32243	33802	4.61	74.59
390	369.87	.948	1582	32816	34398	4.60	75.69
400	388.56	.971	1593	34543	36136	4.41	76.22
410	399.31	.974	1601	35535	37136	4.31	76.60
420	408.40	.972	1605	36376	37981	4.23	76.79
430	424.34	.987	1613	37851	39464	4.09	77.18
440	429.02	.975	1616	38283	39899	4.05	77.32
450	435.48	.968	1619	38881	40500	4.00	77.46
460	445.74	.969	1625	39829	41454	3.92	77.75
470	450.89	.959	1626	40307	41933	3.88	77.80
480	461.58	.962	1631	41296	42927	3.80	78.04
490	471.95	.963	1633	42258	43891	3.72	78.13
500	478.11	.956	1634	42830	44464	3.67	78.18

34 REQUESTS, RUN 13(KWS). TABLE RPL(34,13).

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Output	Overall % Precision	Overall % Known Recall
1	.18	.180	1	5	6	16.67	.33
2	.79	.395	11	16	27	40.74	3.58
3	1.47	.490	19	31	50	38.00	6.19
4	2.18	.545	23	51	74	31.08	7.49
5	2.91	.582	28	71	99	28.28	9.12
6	3.47	.578	33	85	118	27.97	10.75
7	5.26	.751	40	139	179	22.35	13.03
8	5.68	.710	44	149	193	22.80	14.33
9	6.00	.667	44	160	204	21.57	14.33
10	6.00	.600	44	160	204	21.57	14.33
11	7.50	.682	53	202	255	20.78	17.26
12	7.50	.625	53	202	255	20.78	17.26
13	9.50	.731	54	269	323	16.72	17.59
14	10.21	.729	58	289	347	16.71	18.89
15	10.21	.681	58	289	347	16.71	18.89
16	10.21	.638	58	289	347	16.71	18.89
17	11.15	.656	60	319	379	15.83	19.54
18	11.15	.619	60	319	379	15.83	19.54
19	12.82	.675	62	374	436	14.22	20.20
20	13.65	.683	65	399	464	14.01	21.17
21	14.68	.699	67	432	499	13.43	21.82
22	15.79	.718	70	467	537	13.04	22.80
23	16.79	.730	71	500	571	12.43	23.13
24	16.79	.700	71	500	571	12.43	23.13
25	16.79	.672	71	500	571	12.43	23.13
26	16.79	.646	71	500	571	12.43	23.13
27	19.50	.722	71	592	663	10.71	23.13
28	19.50	.696	71	592	663	10.71	23.13
29	20.79	.717	71	636	707	10.04	23.13
30	20.79	.693	71	636	707	10.04	23.13
31	20.79	.671	71	636	707	10.04	23.13
32	20.79	.650	71	636	707	10.04	23.13
33	20.79	.630	71	636	707	10.04	23.13
34	20.79	.611	71	636	707	10.04	23.13
35	20.79	.594	71	636	707	10.04	23.13
36	20.79	.578	71	636	707	10.04	23.13
37	20.79	.562	71	636	707	10.04	23.13
38	20.79	.547	71	636	707	10.04	23.13
39	20.79	.533	71	636	707	10.04	23.13
40	24.65	.616	73	765	838	8.71	23.78
41	24.65	.601	73	765	838	8.71	23.78
42	26.35	.627	75	821	896	8.37	24.43
43	26.35	.613	75	821	896	8.37	24.43
44	26.35	.599	75	821	896	8.37	24.43
45	26.35	.586	75	821	896	8.37	24.43
46	26.35	.573	75	821	896	8.37	24.43
47	26.35	.561	75	821	896	8.37	24.43
48	28.29	.589	76	886	962	7.90	24.76
49	28.29	.577	76	886	962	7.90	24.76
50	28.29	.566	76	886	962	7.90	24.76
51	29.97	.588	77	942	1019	7.56	25.08
52	29.97	.576	77	942	1019	7.56	25.08
53	29.97	.565	77	942	1019	7.56	25.08
54	29.97	.555	77	942	1019	7.56	25.08
55	29.97	.545	77	942	1019	7.56	25.08

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Output	Overall % Precision	Overall % Known Recall
56	29.97	.535	77	942	1019	7.56	25.08
57	29.97	.526	77	942	1019	7.56	25.08
58	29.97	.517	77	942	1019	7.56	25.08
59	29.97	.508	77	942	1019	7.56	25.08
60	29.97	.500	77	942	1019	7.56	25.08
61	29.97	.491	77	942	1019	7.56	25.08
62	29.97	.483	77	942	1019	7.56	25.08
63	29.97	.476	77	942	1019	7.56	25.08
64	29.97	.468	77	942	1019	7.56	25.08
65	29.97	.461	77	942	1019	7.56	25.08
66	29.97	.454	77	942	1019	7.56	25.08
67	33.44	.499	79	1058	1137	6.95	25.73
68	33.44	.492	79	1058	1137	6.95	25.73
69	33.44	.485	79	1058	1137	6.95	25.73
70	33.44	.478	79	1058	1137	6.95	25.73
71	33.44	.471	79	1058	1137	6.95	25.73
72	33.44	.464	79	1058	1137	6.95	25.73
73	33.44	.458	79	1058	1137	6.95	25.73
74	33.44	.452	79	1058	1137	6.95	25.73
75	33.44	.446	79	1058	1137	6.95	25.73
76	35.91	.473	84	1137	1221	6.88	27.36
77	35.91	.466	84	1137	1221	6.88	27.36
78	35.91	.460	84	1137	1221	6.88	27.36
79	35.91	.455	84	1137	1221	6.88	27.36
80	35.91	.449	84	1137	1221	6.88	27.36
81	35.91	.443	84	1137	1221	6.88	27.36
82	39.91	.487	96	1261	1357	7.07	31.27
83	39.91	.481	96	1261	1357	7.07	31.27
84	39.91	.475	96	1261	1357	7.07	31.27
85	44.65	.525	99	1419	1518	6.52	32.25
86	44.65	.519	99	1419	1518	6.52	32.25
87	49.18	.565	101	1571	1672	6.04	32.90
88	49.18	.559	101	1571	1672	6.04	32.90
89	49.18	.553	101	1571	1672	6.04	32.90
90	49.18	.546	101	1571	1672	6.04	32.90
91	49.18	.540	101	1571	1672	6.04	32.90
92	49.18	.535	101	1571	1672	6.04	32.90
93	49.18	.529	101	1571	1672	6.04	32.90
94	49.18	.523	101	1571	1672	6.04	32.90
95	49.18	.518	101	1571	1672	6.04	32.90
96	49.18	.512	101	1571	1672	6.04	32.90
97	49.18	.507	101	1571	1672	6.04	32.90
98	49.18	.502	101	1571	1672	6.04	32.90
99	49.18	.497	101	1571	1672	6.04	32.90
100	49.18	.492	101	1571	1672	6.04	32.90
101	49.18	.487	101	1571	1672	6.04	32.90
102	49.18	.482	101	1571	1672	6.04	32.90
103	49.18	.477	101	1571	1672	6.04	32.90
104	54.79	.527	115	1748	1863	6.17	37.46
105	54.79	.522	115	1748	1863	6.17	37.46
106	54.79	.517	115	1748	1863	6.17	37.46
107	54.79	.512	115	1748	1863	6.17	37.46
108	54.79	.507	115	1748	1863	6.17	37.46
109	54.79	.503	115	1748	1863	6.17	37.46
110	60.74	.552	123	1942	2065	5.96	40.07
111	66.82	.602	124	2148	2272	5.46	40.39
112	73.21	.654	141	2348	2489	5.66	45.93
113	79.06	.700	142	2546	2688	5.28	46.25
114	79.06	.694	142	2546	2688	5.28	46.25
115	79.06	.687	142	2546	2688	5.28	46.25
116	79.06	.682	142	2546	2688	5.28	46.25
117	79.06	.676	142	2546	2688	5.28	46.25
118	84.47	.716	159	2713	2872	5.54	51.79

Cut-off K	Average Output K′	K′/K	Relevant	Irrelevant	Output	Overall % Precision	Overall % Known Recall
119	84.47	.710	159	2713	2872	5.54	51.79
120	84.47	.704	159	2713	2872	5.54	51.79
121	84.47	.698	159	2713	2872	5.54	51.79
122	90.91	.745	163	2928	3091	5.27	53.09
123	90.91	.739	163	2928	3091	5.27	53.09
124	90.91	.733	163	2928	3091	5.27	53.09
125	97.18	.777	164	3140	3304	4.96	53.42
126	97.18	.771	164	3140	3304	4.96	53.42
127	97.18	.765	164	3140	3304	4.96	53.42
128	97.18	.759	164	3140	3304	4.96	53.42
129	103.88	.805	171	3361	3532	4.84	55.70
130	103.88	.799	171	3361	3532	4.84	55.70
131	103.88	.793	171	3361	3532	4.84	55.70
132	103.88	.787	171	3361	3532	4.84	55.70
133	103.88	.781	171	3361	3532	4.84	55.70
134	103.88	.775	171	3361	3532	4.84	55.70
135	103.88	.769	171	3361	3532	4.84	55.70
136	103.88	.764	171	3361	3532	4.84	55.70
137	103.88	.758	171	3361	3532	4.84	55.70
138	103.88	.753	171	3361	3532	4.84	55.70
139	110.53	.795	183	3575	3758	4.87	59.61
140	117.62	.840	187	3812	3999	4.68	60.91
141	117.62	.834	187	3812	3999	4.68	60.91
142	117.62	.828	187	3812	3999	4.68	60.91
143	117.62	.823	187	3812	3999	4.68	60.91
144	117.62	.817	187	3812	3999	4.68	60.91
145	117.62	.811	187	3812	3999	4.68	60.91
146	117.62	.806	187	3812	3999	4.68	60.91
147	117.62	.800	187	3812	3999	4.68	60.91
148	117.62	.795	187	3812	3999	4.68	60.91
149	117.62	.789	187	3812	3999	4.68	60.91
150	117.62	.784	187	3812	3999	4.68	60.91
151	117.62	.779	187	3812	3999	4.68	60.91
152	117.62	.774	187	3812	3999	4.68	60.91
153	117.62	.769	187	3812	3999	4.68	60.91
154	117.62	.764	187	3812	3999	4.68	60.91
155	117.62	.759	187	3812	3999	4.68	60.91
156	117.62	.754	187	3812	3999	4.68	60.91
157	117.62	.749	187	3812	3999	4.68	60.91
158	117.62	.744	187	3812	3999	4.68	60.91
159	117.62	.740	187	3812	3999	4.68	60.91
160	117.62	.735	187	3812	3999	4.68	60.91
161	117.62	.731	187	3812	3999	4.68	60.91
162	117.62	.726	187	3812	3999	4.68	60.91
163	117.62	.722	187	3812	3999	4.68	60.91
164	117.62	.717	187	3812	3999	4.68	60.91
165	117.62	.713	187	3812	3999	4.68	60.91
166	117.62	.709	187	3812	3999	4.68	60.91
167	117.62	.704	187	3812	3999	4.68	60.91
168	123.32	.734	188	4005	4193	4.48	61.24
169	123.32	.730	188	4005	4193	4.48	61.24
170	123.32	.725	188	4005	4193	4.48	61.24
171	123.32	.721	188	4005	4193	4.48	61.24
172	123.32	.717	188	4005	4193	4.48	61.24
173	123.32	.713	188	4005	4193	4.48	61.24
174	123.32	.709	188	4005	4193	4.48	61.24
175	123.32	.705	188	4005	4193	4.48	61.24
176	131.65	.748	191	4285	4476	4.27	62.21
177	131.65	.744	191	4285	4476	4.27	62.21
178	131.65	.740	191	4285	4476	4.27	62.21
179	131.65	.735	191	4285	4476	4.27	62.21
180	131.65	.731	191	4285	4476	4.27	62.21
181	131.65	.727	191	4285	4476	4.27	62.21
182	131.65	.723	191	4285	4476	4.27	62.21

Cut-off K	Average Output K'	K'/K	Rele-vant	Irrel-evant	Output	Overall % Prec-ision	Overall % Known Recall
183	140.53	.768	194	4584	4778	4.06	63.19
184	140.53	.764	194	4584	4778	4.06	63.19
185	140.53	.760	194	4584	4778	4.06	63.19
186	140.53	.756	194	4584	4778	4.06	63.19
187	140.53	.751	194	4584	4778	4.06	63.19
188	140.53	.748	194	4584	4778	4.06	63.19
189	140.53	.744	194	4584	4778	4.06	63.19
190	150.56	.792	201	4918	5119	3.93	65.47
191	150.56	.788	201	4918	5119	3.93	65.47
192	150.56	.784	201	4918	5119	3.93	65.47
193	150.56	.780	201	4918	5119	3.93	65.47
194	161.24	.831	204	5278	5482	3.72	66.45
195	169.88	.871	207	5569	5776	3.58	67.43
196	169.88	.867	207	5569	5776	3.58	67.43
197	169.88	.862	207	5569	5776	3.58	67.43
198	180.09	.910	213	5910	6123	3.48	69.38
199	180.09	.905	213	5910	6123	3.48	69.38
200	180.09	.900	213	5910	6123	3.48	69.38
210	190.00	.905	213	6247	6460	3.30	69.38
220	197.71	.899	213	6509	6722	3.17	69.38
230	197.71	.860	213	6509	6722	3.17	69.38
240	211.03	.879	221	6954	7175	3.08	71.99
250	220.88	.884	223	7287	7510	2.97	72.64
260	220.88	.850	223	7287	7510	2.97	72.64
270	229.44	.850	228	7573	7801	2.92	74.27
280	242.74	.867	229	8024	8253	2.77	74.59
290	242.74	.837	229	8024	8253	2.77	74.59
300	242.74	.809	229	8024	8253	2.77	74.59
310	286.09	.923	237	9490	9727	2.44	77.20
320	302.09	.944	238	10033	10271	2.32	77.52
330	302.09	.915	238	10033	10271	2.32	77.52
340	302.09	.889	238	10033	10271	2.32	77.52
350	302.09	.863	238	10033	10271	2.32	77.52
360	302.09	.839	238	10033	10271	2.32	77.52
370	302.09	.816	238	10033	10271	2.32	77.52
380	302.09	.795	238	10033	10271	2.32	77.52
390	302.09	.775	238	10033	10271	2.32	77.52
400	302.09	.755	238	10033	10271	2.32	77.52
410	302.09	.737	238	10033	10271	2.32	77.52
420	302.09	.719	238	10033	10271	2.32	77.52
430	302.09	.703	238	10033	10271	2.32	77.52
440	302.09	.687	238	10033	10271	2.32	77.52
450	326.15	.725	241	10848	11089	2.17	78.50
460	326.15	.709	241	10848	11089	2.17	78.50
470	344.00	.732	243	11453	11696	2.08	79.15
480	344.00	.717	243	11453	11696	2.08	79.15
490	344.00	.702	243	11453	11696	2.08	79.15
500	344.00	.688	243	11453	11696	2.08	79.15

APPENDIX B9

93 Requests: Numbers of Relevant Documents Retrieved

with (in brackets) the numbers of requests making a positive contribution to each total

Run	K′ (nearest integer) 20	30	40	45	46	47	48	49	50
KWS 13	528(85)	671(85)	897(87)	928(87)	954(87)	961(87)	961(87)	991(87)	991(87
AWKWS 22	499(88)	650(88)	781(88)	850(88)	876(88)	890(88)	890(88)	898(88)	898(88
ARM 16	402(78)	525(81)	656(83)	773(84)	773(84)	773(84)	781(85)	781(85)	782(86
MCS01 14	394(78)	549(81)	753(82)	812(83)	829(83)	834(83)	844(83)	849(83)	873(85
MCS11 20	428(81)	630(87)	752(87)	813(87)	820(88)	828(88)	828(88)	857(88)	857(88
RJR 19	489(82)	611(84)	739(86)	807(86)	848(86)	860(86)	865(86)	880(86)	939(86
ARMSR 17	401(78)	537(82)	637(83)	687(85)	711(85)	775(85)	775(85)	785(86)	785(86
SR14 6	395(83)	539(89)	649(91)	706(91)	719(91)	735(91)	746(91)	753(91)	753(91
PDR14 11	425(83)	551(87)	669(90)	776(90)	781(90)	804(90)	820(90)	831(90)	861(90
EAG3 15	405(77)	502(83)	660(85)	704(86)	785(87)	785(87)	800(87)	828(87)	828(87
EAG4 18	405(84)	541(86)	725(89)	815(89)	815(89)	819(89)	819(89)	835(90)	835(90
EARG4 23	460(81)	613(86)	783(88)	845(88)	871(88)	871(88)	888(89)	895(89)	914(89
13T14 9	486(85)	632(87)	848(87)	884(87)	894(88)	910(88)	910(88)	910(88)	966(88
13W14 28	526	629	860	*917	*920	*930	*939	*949	*991(89
U14 21	473(80)	602(82)	782(83)	840(84)	842(84)	867(84)	868(84)	880(84)	902(84
Union of Standard Strategies	1195(92)	1528(92)	1816(93)	1900(93)	1925(93)	1944(93)	1954(93)	1976(93)	2020(93

APPENDIX B9a

93 Requests: Numbers of NEW Relevant Documents Retrieved,

that is, not also retrieved by Run 13(KWS) at same K′,

with (in brackets) the numbers of requests making a positive contribution to each total

Run	K′ (nearest integer) 20	30	40	45	46	47	48	49	50
KWS 13	-	-	-	-	-	-	-	-	-
AWKWS 22	105(32)	132(36)	117(32)	130(33)	115(32)	113(32)	113(32)	108(33)	108(33)
ARM 16	121(41)	144(50)	185(51)	210(52)	209(51)	207(51)	211(53)	206(53)	206(53)
MCS01 14	130(43)	184(49)	207(51)	221(55)	227(55)	225(56)	233(57)	234(57)	252(58)
MCS11 20	104(41)	179(53)	157(50)	189(50)	192(51)	197(50)	197(50)	196(51)	196(51)
RJR 19	115(39)	147(39)	162(41)	160(41)	152(42)	156(42)	157(43)	161(44)	218(44)
ARMSR 17	116(41)	152(50)	171(49)	194(50)	205(50)	207(51)	207(51)	206(53)	206(53)
SR14 6	145(51)	203(59)	213(68)	222(69)	219(68)	233(68)	238(68)	238(67)	238(67)
PDR14 11	140(44)	179(53)	178(65)	230(67)	223(67)	235(68)	242(69)	245(68)	271(69)
EAG3 15	152(47)	168(53)	241(58)	256(59)	291(60)	287(60)	295(60)	306(60)	306(60)
EAG4 18	129(49)	184(56)	275(55)	299(58)	297(57)	296(59)	296(59)	290(59)	290(59)
EARG4 23	157(47)	210(59)	212(63)	227(64)	232(63)	228(63)	240(64)	234(63)	246(63)
13T14 9	101(32)	131(37)	126(37)	125(37)	121(36)	124(39)	124(39)	122(37)	176(42)
13W14 28									*155(32)
U14 21	200(49)	234(50)	257(53)	283(58)	272(58)	[illegible]2(58)	293(57)	296(57)	311(58)
Union of Standard Strategies	667(86)	857(88)	919(90)	972(89)	971(89)	983(89)	993(89)	985(89)	1029(89)

*output not wholly assessed

APPENDIX B10

34 Requests Retrieving 0 to 4 Relevant Documents in Run 13 (KWS) at K=71, $K' \simeq 50$

Request	Known Relevant	Relevant in Run 13	Output in Run 13	Next Highest Output containing more relevant	P = less well formulated
3	15	4	14	242	
5	5	4	125	1568	P
7	4	0	12	375	
11	1	1	70	-	
12	2	2	79	-	P
13	11	4	19	360	P
14	4	0	51	1884	
35	11	3	47	341	
38	5	4	70	264	
40	8	1	24	371	
43	5	2	32	850	P
49	7	4	34	317	
50	8	3	4	165	
51	5	1	9	163	
54	3	3	38	-	P
55	7	4	39	583	P
56	1	0	7	214	
60	10	1	8	210	P
64	10	3	74	539	
65	1	0	50	502	
67	3	1	80	415	
71	22	0	25	209	
72	4	2	18	231	
73	8	4	19	260	
74	20	4	25	251	
77	5	4	40	573	
78	21	3	8	199	
81	32	3	3	220	
87	13	4	11	464	
88	4	1	31	333	
92	8	2	12	231	
94	7	3	13	212	P
96	12	0	13	149	P
99	14	4	33	117	
Totals	296	79	1137	12812 (31 requests)	
Averages	8.7	2.3	33.41	413	

Of the above, and omitting less well formulated requests, 7,14,35,38,40,49,64,65,67,73,77,87,88 have their next highest <u>productive</u> output greater than 250.

APPENDIX B11

Subtotals for 34 Requests, Standard Runs at $K' \simeq 50$

With (in brackets) numbers of requests making a positive contribution

New=Not also retrieved in Run 13

Run	Relevant	Irrelevant	Unassessed	Output	Effective K′	New Relevant	New Irrelevant
KWS 13	79(28)	1058	0	1137	33	-	-
AWKWS 22	95(29)	1467	0	1562	46	17(8)	508
ARM 16	108(30)	1640	0	1748	51	39(15)	906
MCSO1 14	109(28)	1578	0	1687	50	59(16)	1038
MCS11 20	111(30)	1428	0	1539	45	46(14)	770
RJR 19	101(28)	1082	0	1193	35	33(9)	440
ARMSR 17	109(29)	1643	0	1752	52	40(15)	928
SR14 6	109(32)	1537	0	1646	48	49(21)	1011
PDR14 11	136(31)	1539	0	1675	49	69(22)	987
EAG3 15	141(30)	1464	0	1605	47	77(20)	958
EAG4 18	139(32)	1504	0	1643	48	77(23)	1035
EARG4 23	132(30)	1498	0	1630	48	70(21)	884
13T14 9	134(29)	1632	0	1766	52	64(17)	868
13W14 28	130(30)	1483	30	1643	48	54(12)	607
U14 21	122(27)	1451	0	1573	46	78(19)	1109
Union of Standard Strategies	296(34)	5835	0	6131	-	217(31)	4777

APPENDIX B12

Request sets of increasing generality (report, VIII.4)

Run	Numbers of Relevant Documents at K' ≃ 50 GEN 1	GEN 2	GEN 3	GEN 4	Numbers of New Relevant Documents GEN 1	GEN 2	GEN 3	GEN 4
KWS 13	68(22)	137(20)	277(22)	509(23)	-	-	-	-
AWKWS 22	72(22)	128(20)	285(23)	413(23)	11 (6)	13 (5)	37(11)	47(11)
ARM 16	62(23)	102(20)	199(22)	419(21)	14 (9)	20(11)	44(14)	128(19)
MCS01 14	50(21)	104(20)	238(21)	454(21)	12 (7)	29(14)	57(17)	154(20)
MCS11 20	64(23)	118(20)	232(22)	443(23)	12 (7)	34(10)	46(15)	104(19)
RJR 19	56(20)	130(21)	258(22)	495(23)	4 (2)	33(13)	47(16)	134(13)
ARMSR 17	63	101	205	416	15	20	44	127
SR14 6	70(24)	111(21)	225(23)	347(23)	21(14)	27(12)	72(21)	118(20)
PDR14 11	74(23)	124(21)	272(23)	391(23)	24(13)	37(15)	85(19)	125(22)
EAG3 15	57(22)	123(20)	239(23)	409(22)	16 (8)	40(14)	85(17)	165(21)
EAG4 18	69(24)	121(21)	212(23)	433(22)	22(12)	39(13)	54(15)	175(19)
EARG4 23	58(23)	121(21)	243(23)	492(23)	14(11)	34(11)	74(21)	124(20)
13T14 9	64(21)	141(21)	258(23)	503(23)	11 (8)	32(11)	50(10)	83(13)
13W14 28	64(22)	146(21)	277(23)	504(23)	3 (3)	29 (9)	45 (9)	78(11)
U14 21	55(21)	136(20)	257(21)	454(22)	19 (9)	48(16)	73(15)	171(18)
Union of Standard Strategies	136(26)	262(21)	539(23)	1083(23)	68(23)	125(20)	262(23)	574(23)

APPENDIX B13

Subtotals for 55 Requests, Runs with Underlining and other Runs all at $K' \simeq 30$

$\overline{25}$ = Not also retrieved in Run 25(U13)

Run		Relevant	Irrelevant	Unassessed	Output	Effective K'	Relevant $\overline{25}$	Irrelevant $\overline{25}$
KWS	13	448	1288	0	1736	32	152	772
MCS01	14	390	1271	0	1661	30	180	960
PDR14	11	392	1269	0	1661	30	176	985
U13	25	478	1130	14	1622	29	-	-
U14	21	419	1049	0	1468	27	64	330
U11	24	412	950	41	1403	26	60	325
AWKWS	22	432	1356	0	1788	33	124	919
ARM	16	340	1224	0	1564	28	148	996
MCS11	20	425	1308	0	1733	32	192	955
RJR	19	418	1201	0	1619	29	158	846
ARMSR	17	353	1248	0	1601	29	153	1020
SR14	6	382	1272	0	1654	30	188	1010
EAG3	15	336	1155	0	1491	27	156	937
EAG4	18	336	1179	0	1515	28	157	982
EARG4	23	431	1252	0	1683	31	217	984
13T14	9	445	1261	0	1706	31	181	872
Union of Standard Strategies		1065	5345	54	6464	-	587	4215

APPENDIX B14

The Approximate value of $\frac{K'}{K}$.

Take the single request in V.6.3 with its choice of output quantities, 0, 3, 7, 15, ,..., thus

K	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	...
output	0	0	3	3	3	5	7	7	7	7	7	11	15	15	15	15	15	15	15	...
ΣK	0	1	3	6	10	15	21	28	36	45	55	66	78	91	105	120	136	153	171	...
Σ (output)	0	0	3	6	9	14	21	28	35	42	49	60	75	90	105	120	135	150	165	...

where for the moment we have written an average in the case of a tie (K = 5,11) instead of going to the lower (3,7) ...(A). In the third and fourth rows we have accumulated these values. Note how Σ (output) lags behind ΣK, catching up only at the points where K or K+1 is an exact output. In fact

$$\sum_{1}^{K} (\text{output}) = \tfrac{1}{2}K(K+1) - \tfrac{1}{2}(K-y)(K-y+1)$$

where y is the output nearest to K, whether above or below.

We now use the assumption that for a <u>set of requests</u> $\frac{K'}{K}$ is approximately constant, as K varies. In particular

$$\frac{K'}{K} \simeq \frac{\sum_{1}^{K} K'}{\sum K} = \frac{\frac{1}{n}\sum_{\text{req}}\sum_{1}^{K} (\text{output})}{\sum K} \quad (n \text{ requests})$$

$$= \frac{\sum_{\text{req}} \{K(K+1) - (K-y)(K-y+1)\}}{nK(K+1)} \qquad \text{(B)}$$

$$\simeq 1 - \frac{1}{n}\sum_{\text{req}} \left(1 - \frac{y}{K}\right)^2 ,$$

y varying from request to request.

Thus the larger the spread of the y values, the further below 1 is $\frac{K'}{K}$.

We can take (B) further. The y values have mean K', and standard deviation vK', say. Then

$$\sum_{req} y = nK'$$

$$\sum_{req} y^2 = nK'^2(1+v^2).$$

From (B)

$$\frac{K'}{K} \simeq \frac{\sum_{req}(2Ky-y^2+y)}{nK(K+1)} = \frac{2KK'-K'^2(1+v^2)+K'}{K^2+K}$$

which yields the simple relation

$$\frac{K'}{K} \simeq \frac{1}{1+v^2},$$

where $v = \frac{\text{s.d.}}{\text{mean}}$ for the outputs whose average is K'.

The adjustment we made in (A) affects the final result by at most 1 part in 2K and may be ignored.

More generally, if the outputs have the same dispersion or coefficient of variation v at $K = K_1$ and $K = K_2$ and $\frac{K'}{K}$ may be taken as constant from K_1 to K_2, we may still deduce in a similar manner that

$$\frac{K'}{K} \simeq \frac{1}{1+v^2}.$$

v will be considerable for a strategy in which the ratio r between successive output totals tends to be large. Take an imaginary case in which r is constant for all requests and all levels and abandon the stipulation that the y's must be integers. $\frac{K'}{K}$ being constant puts a second constraint on the y values nearest to a given K. If, e.g., no two are equal, a little experiment shows that they must be in geometric progression with ratio $r^{\frac{1}{n}}$, say

$$z,\ zr^{\frac{1}{n}},\ zr^{\frac{2}{n}},\ \ldots\ldots,\ zr^{\frac{n-1}{n}}$$

where e.g. K is just nearer z than zr but just near $zr^{\frac{n-1}{n}}$ than $zr^{\frac{1}{n}}$. $r^{\frac{1}{n}}$ will be close to 1.

Thus $$K \simeq \frac{1}{2}(z+zr)$$

$$K' = \frac{1}{n}\left(z + zr^{\frac{1}{n}} + \ldots + zr^{\frac{n-1}{n}} \right)$$

$$= \frac{1}{n} . z . \frac{r-1}{r^{\frac{1}{n}} - 1} \simeq z . \frac{r-1}{\log_e r}$$

and $$\frac{K'}{K} \simeq 2 . \frac{r-1}{r+1} . \frac{1}{\log_e r} .$$

Thus if $r = 10$, $\frac{K'}{K} \simeq .76$, which could be typical values for run 13. If $r = 2$, $\frac{K'}{K} = \frac{2}{3 \times .69} = .96$, while if $r = 3$, $\frac{K'}{K} = .91$ and the effect is becoming noticeable.

APPENDIX B15

Words Accepted by Assessors in Place of Key-Words

A short study was made by S. Whelan of abstracts retrieved by Subject Indexes which had not come up in any mechanical strategy. In particular, these documents would not have had sufficient keywords for retrieval in Run 13 at $K' \simeq 51$.

For example:-

Request 74. Predicting the paths of electrons moving in a varying magnetic field.

Abstract 5163. The motion of charged particles in weakly variable magnetic fields. Analysis is simplified by considering in place of the actual particle which follows a helical path an equivalent particle following a mean path and having a magnetic moment.

Abstract 7855. On the non-optical theory of focusing in rotating magnetic fields. The difficulties arising in the application of optical methods to electron optics are pointed out. The general theory of the focusing action of static magnetic fields developed by Grinberg which is based on nonoptical methods is applied to the case of apraxial electron beams in rotating magnetic fields.

Here we have 'charged particle' accepted by the requester in place of 'electron' while 'focusing' corresponds to the word 'paths'. For want of a better term, let us call these words or phrases 'acceptances'. These acceptances were examined to see whether they were in G3, and below are some that were not.

Abstract Word	Request Word	Subjective Judgement on Preserving Connection in a Word-Word Matrix
Charge	Electron	Yes
Current	Electron	Yes
Density	Ion	Yes
Discharge	Emission	?
Films	Oxidation	No
Focus	Path	Yes
Impact	Bombardment	Yes
Illumination	Bombardment	?
Motion	Moving	Yes
Particle	Electron	Yes
Production	Emission	No
Pulse	Electron	?
Screen	Cathode	Yes
Surface	Cathode	?
Trajectory	Path	Yes
Trajectory	Electron	?
Velocity	Bombardment	No
Waveform	Spectrum	Yes

We reproduce the rows of G3 corresponding to some of the stems on the request and abstracts quoted:-

Stem	Stems Associated in G3					
charg(e)	particle	relativi	trajecto	charg	space	mass
electron	attachme	collisio	electron	nitrogen	secondar	transpor
	neutral	profile	diffus	target	valenc	gases
	probe	atom	ion			
field	anisotro	corrugat	homogene	magnetor	strength	transpor
	transver	classic	moment	tensor		
foc(us)	foc					
magnetic	ferromag	magnetic	magnetis	ferrite	magneti	saturat
	moment	fluid	force	gauss	drum	head
motion	perturb	motion				
moving	Cerenkov	particle	relativi	velocit	moving	
particle	particle	relativi	trajecto	moving	proton	charg
	force	outer	belt	trap		
path	path					
predict	predict					
trajecto(ry)	particle	trajecto	charg			
vary	vary					

Examination of such lists as the above led to the following simple observations:-

1. If it were desired to amend or generate a word-word thesaurus by this means, we would need many more than 93 requests, and new criteria of frequency of association.

2. G3 is, after all, generated from the document set as a whole, and any threshold must exclude some connections which are only 'part of the picture'.

3. The particular case of 'motion-moving' arises because the stem mo- is too short to be used in the basic vocabulary. Where grammatical variants cannot be united by a common stem, they should be by some other means, for example, by assigning the variant stems the same code number in the dictionary, and choosing one as the preferred term for ease of reference. If they were assigned different code-numbers and links were inserted in the word-word matrix, joining each to the other and its associates, this would be less straightforward and would not help key-word-stem strategies.

APPENDIX B16

Sample Page of Subject Indexes to Abstracts

(see VII.5)

Electron microscopy, application at A.E.I. Research Laboratory, **2740**
at Cambridge, **2740**
conference in London, Nov. 1953, **3330**
French work in during 1952–1953, **2740**
replicas for, high-resolution, preparation of, **3328**
shadowing technique for, **3653**
Swiss work in, **2740**
Electron motion, in electric and magnetic fields, **105, 1384**
Electron multipliers. See Photocells and Valves.
Electron optics, double focusing by two-magnet system, **498**
focusing charged particles, **3541**
magnetic and electric fields with cylindrical and mirror symmetry, **2191**
relativistic aberration functions in, **3649**
schlieren technique for studying e.m. fields, **1536**
shadow method of mapping magnetic fields, **2465**
wave-mechanics theory of, **499**
Electron spectrographs, e.s., developed from electron microscope, **3655**
Electron trajectories, automatic plotting of, use of electrolyte tank for, **3304**
numerical integrations of equation of, **1535**
Electronic applications, (See also Control systems; Heating; Photocells, applications of)
aircrew training equipment using analogue computers, **3477**; air-warfare game, **2187**; blind guiding device, **3641**; electrocardiology, resultant dipole of heart, **2735**; flying-spot scanner for studying visual perception, **1541**; metal detector, simplified construction, ASDET, **3315**; microwave technique for observing commutator surfaces during operation, **2736**; monitoring high-speed phenomena, **3319**; nondestructive testing of materials, **3313**; in oil industry, **3660**; photoelectric device for indicating position of rotating shaft, **812**; pressure measurement in i.c.e., **3332**; short-time-interval measurement using two photocells, **3617**; tachometer, using cold-cathode tubes for visual display, **1153**; thermostat, with variable-duty-factor heating cycle, **3312**; vibration meter, self-contained, **3309**; vibration methods of testing, **1881**
Electronic engineering, developments in, **1267**
Electronic equipment, for aircraft, reliability of, improvement of, **609**
automatic assembly of, **694**
construction of, modern techniques for, **101**
manufacture of, importance of physics in, **897**
modular design and mechanized production in, **3432**
mechanized production of, N.B.S. methods for, **693**
temperature calculations for, **291**
Electrons, classical theory of, **2066, 3188**
and crystal-lattice vibrations, model for investigating interaction between, **108, 3517**
energy loss in passage through thin films, **2352**
high-energy, deflection in magnetized iron, **1041**
new classical theory of, **3187**
slow, effect of bombardment of thin films by, **1748**
motion of in air, with application to ionosphere, **106**
Electrophoresis, insulation of transformer-core tapes by, **3612**
in valve manufacture, **3720**
Electrostatics, field and induction of point charge in space, **2353**
Equalization, (See also Networks; Transmission lines)
comparison of time-function and frequency-function methods for, **2598**
Exhibitions, British Instrument Industries, London, 1953, **608**; components, Paris, 1954, **3426**; German Industries, Hanover, 1954, **3427**; German radio, Düsseldorf, 1953, **292**; Hanover Technical Fair, 1953, electrical measurement equipment at, **208**, radio, television and electroacoustic apparatus at, **293**; Leipzig fair, 1953, **610**; Lyons fair, 1954, Soviet radio equipment at, **3431**; Physical Society, 1954, **2272**; R.E.C.M.F., London, April 1954, **2063**; S.B.A.C., Farnborough, 1954, **3734**; 16th Salon national, Paris, 1953, radio and television at, **1171**; 20th Salon de l'Aéronautique, 1953, **1268**; 21st National Radio, London, 1954, **3735**

Faraday effect, cm-λ, experimental investigation of, **3202**
in ferrites, **995**
in waveguides and cavity resonators, **2987**
at microwave frequencies, medium-boundary effects in, **2927**
in waveguides containing anisotropic media, **1300**
Ferrimagnetic materials, ferrites, after-effects in, **3605**
as core materials for magnetic heads, **2846**
of Dy and Er, thermomagnetic study of, **3274**
fermalite, fernilite and fercolite, **1827**
ferromagnetic resonance of, **1832**, effect of anisotropy and relaxation phenomena on, **2713**
ferroxcube, measurement of properties of, **207**
ferroxcube IV, thermal effects accompanying magnetization of, **177**
ferroxdure, $BaO.6Fe_2O_3$, with preferred orientation, **184**, domain-wall displacements in, **179**
Gd, thermomagnetic study of, **3273**
magnetic phenomena in, **175**
magnetic resonance phenomena in, **2144**
magnetostrictive, as u.h.f. detectors, **2904**
Mg, dielectric behaviour of, **1107**, investigation by heat treatment, **1108**

Ferrimagnetic materials, ferrites, MgAlMn, having low loss at 4 kMc/s, **3607**
Mg-Mn, commercially available, **2146**
microwave Faraday effect in, **995**
mixed, cation distributions in, **2986**
mixed crystals of, spontaneous magnetization of, calculation of, **3606**
Mn, Mg and Co, magnetic susceptibility of, **2443**
modern developments in, **178, 1823**
Ni, mixed with nickel titanate, magnetic properties of, **2711**
Ni-Fe, temperature dependence of resonance line width in, **2446**
Ni-Zn, magnetic-dispersion spectrum of, **2710**, temperature dependence of magnetization of, **1488**
nickel ferrite aluminate, g factor of, **2985**, resonance absorption in, **2145**
nickel ferrite-aluminates and -gallates, **1830**
properties and applications, **1487, 3608**
rare-earth, influence of ionic diameters on properties of, **1109**, molecular theory of, **3275**
relaxation of permeability in, effect of dimensions and compression on, **1829**
relaxation phenomena in, **2445**
resonance absorption in, **2712**
magnetic resonance in **1828**
measurement of permeability tensor in, **769**
spinels, ionic distribution deduced from g-factor of, **1831**
Ferroelectric materials, $BaTiO_3$, colloidal, properties of, **3244**
double hysteresis loop of at Curie temperature, **195**
effect of two-dimensional pressure on, **2116**
temperature changes of properties of, **1116**
variation of hysteresis loop with age, **1117**
$BaTiO_3$ single crystals, dielectric properties of, **751**
domain clamping effect in, **3245**
$BaTiO_3$-$PbZrO_3$ solid solutions, experimental determination of, **1446**
cadmium and lead niobates, **1072, 1073**
ceramics, coefficients of, methods for measurement of, **2720**
effects of aging on, **1444**
colloidal, theory of, **1445**
domain-boundary movements in, possible mechanism of, **3243**
KH_2PO_4, colloidal, properties of, **776**
$KNbO_3$, third phase transition in, **2115**
$Pb(NbO_3)_2$, properties of, **1448**
$PbZrO_3$, antiferroelectric-ferroelectric phase transition in, **3246**
$PbZrO_3$ and $PbTiO_3$, solid solutions of, **1121**
perovskite-type mixed-crystal titanates, properties of, **3572**
Rochelle salt, experimental investigation of, **2965**
theory of, **2964**
titanates, experimental investigation of, **194**
solid solutions of, phase transitions in, **3278**
Ferroelectricity, in $BaTiO_3$ and other perovskite-type crystals, and crystal structure, **2410**
origin of, **1789**
Ferromagnetic materials, alloys, with α-Fe lattice, after-effect in, **1822**
binary, saturation magnetization of, theory of, **1105**
Cr-Te, ferromagnetic resonance in, **180**
Gd-Mg, **185**
Ni-Fe, magnetic viscosity of, temperature dependence of, **181**, oriented, recent progress in manufacture of, **453**
Ni-Mn, effect of composition on change of resistance in magnetic field, **1820**
precipitation, magnetic viscosity in, **3272**
Pt-Co, magnetic viscosity of, **767**
Si-Fe, for communication applications, **(A)771**, grain-oriented and single-crystal, powder patterns of, **770**, preparation of large single crystals of, **182**, single crystals, domain processes in, **451**
16% Al-Fe, cold-rolled at 575°C, **2456**
armco iron, influence of domain thickness on permeability of, **3271**
with large domains, permeability at 0–7000 Mc/s of, **2984**
cementite (Fe_3C), experimental investigation of, **1104**
for cores, assessment of variability of, **183**
evaluation of, **466**
Curie points of, change of with pressure, **2142**
diffusion after-effect in, variation with induction, **3197**
in weak alternating fields, **2637**
domains in, observation of by longitudinal Kerr effect method, **2441**
Fe, thin films of, magnetization of, **2709**
Fe and Ni, Curie temperatures of, effect of pressure on, **2143**
effect of plastic deformation on magnetization of, **1835**
ferrites. See Ferrimagnetic materials.
hard, theory of hysteresis in, **2141**
iron amalgam, effect of grain size on magnetic properties of, **449**
irons and steels, permeability of, dependence on mechanical stress of, **2440**
irreversible after-effect in, **2074**
magnetic viscosity of, **176**
magnetization curves for single crystals, influence of domain structure on, **117**
magnetostriction in, expressed in terms of external field, **2988, 3276**
manganites, experimental investigation of, **3604**
maximum-stability state in, definition of, **1102**
metals, apparent permeabilities at cm λ of, experimental determination of, **1505**
metals and alloys, magnetic moments and crystal structure of, **1484**
temperature dependence of spontaneous magnetization of, **3609**
mild steel, effects of tension and compression on magnetic characteristics of, **446**

APPENDIX B17

Glossary

association factor	III.6
coverage	IX.1; VIII.2
coordination	I.13, appendix B4
descriptor	VI.2.2
document = abstract, both including title	
G3, G4	III.10
K (desired output cut-off)	V.6
K' (average output)	V.6
relevant	VII.1
sensitivity	V.7, VIII.2, IX.1
similarity coefficient	III.9
strategy = run = retrieval method = formula for coordination	
stratum	V.6.

BIBLIOGRAPHY

1. GARDIN, J.C. Le Syntol (syntagmatic organization language). Etude d'un systeme general de documentation automatique. Assoc. Marc Bloch, Final Report to Euratom Part 1, 1962.

2. FARRADANE, J. Relational indexing and classification in the light of recent experimental work in psychology. Information Storage and Retrieval, 1963, 1, (1), 3.

3. KENT, A.K. et al. Chemical Society Research Unit, Nottingham University, Bulletin, Jan., 1969.

4. VASWANI, P.K.T. Mechanized storage and retrieval of information. Rev. int. Docum., 1965, 32, (1), 19.

5. MEETHAM, A.R. Preliminary studies for machine generated index vocabularies. Language and Speech, 1963, 6, (1), 22.

6. STILES, H.E. The association factor in information retrieval. J. Assoc. Comput. Mach., 1961, 8, (2), 271.

7. Statistical association methods for mechanized documentation, Washington, 1964, edited by V.E. Giuliano, N.B.S. Misc. Pub. 269.

8. MEETHAM, A.R. Probabilistic pairs and groups of words in a text. Language and Speech, 1964, 7, (2), 98.

9. VASWANI, P.K.T. A technique for cluster emphasis and its application to automatic indexing. Proc. IFIP Congress, Edinburgh, 1968, edited by A.J.H. Morrell, Amsterdam, North-Holland Publishing Co., 1969, 2, 1300.

10. MEETHAM, A.R. Graph separability and word grouping. Proc. 21st National Conference of A.C.M., (1966), London, Academic Press, 1967, 513-4.

11. MEETHAM, A.R. Algorithm to assist in finding the complete sub-graphs of a given graph. Nature, 1966, 211, (5044), 105.

12. SPARCK JONES, K. and JACKSON, D. Current approaches to classification and clump-finding at the Cambridge Language Research Unit. Computer J., 1967, 10 (1), 29.

13. SPARCK JONES, K. and JACKSON, D. The use of automatically-obtained keyword classifications for information retrieval. Final report, 1969, ML211, Cambridge Language Research Unit, Cambridge.

14. BAR-HILLEL, Y. Some theoretical aspects of the mechanization of literature searching. Technical report no. 3, Hebrew University, Jerusalem, 1960.

15. CLEVERDON, C., MILLS, J. and KEEN, M. Factors determining the performance of indexing systems. 2 vols., 1966, College of Aeronautics, Cranfield.

16. MOOERS, C.N. Zatocoding applied to mechanical organization of knowledge. American Documentation, 1951, 2, 20.

17. GIULIANO, V.E. The interpretation of word associations. Page 26 in ref. 7 above.

18. BORKO, H. The construction of an empirically based mathematically derived classification system. Proc. Spring Joint Computer Conference, San Fransisco, 1962, AFIPS, 1962, 21, 279.

19. MARON, M.E. and KUHNS, J.L. On relevance, probabilistic indexing and information retrieval. J. Assoc. Comput. Mach., 1960, 7, (3), 216. (One of the earliest papers suggesting the use of association measures).

20. STEVENS, M.E. Automatic indexing: a state-of-the-art report. Nat. Bur. Stand. Monograph 91, Washington, 1965.

21. YULE, G.U. and KENDALL, M.G. An introduction to the theory of statistics, 14th edition, Charles Griffin & Co. Ltd., London, 1950.

22. WOLFBERG, M.S. Determination of maximally complete subgraphs. Moore School Report No. 65-27, 1965, University of Pennsylvania, Philadelphia, Pa.

BM